Polyglutamine-mediated and Heavy Metal-induced Transcription from Yeast to Humans

---

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde

(Dr. sc. nat)

vorgelegt der Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Lilit Atanesyan

aus

Armenien

Promotionskomitee

Prof. Dr. Walter Schaffner (Vorsitz und Leiter der Dissertation, UZH, Zürich)

Prof. Dr. Konrad Basler (UZH, Zürich)

Prof. Dr. Henrik Kaessmann (UNIL, Lausanne)

Zürich, 2011

TABLE OF CONTENTS

Summary

During my PhD thesis I was following two major projects, namely, the effect of polyglutamine tracts on the activity of a specific transcription factor, and on regulation of heavy metal homeostasis and detoxification.

Project I: Many proteins, especially regulatory proteins of gene expression, contain homopolymeric repeats of single amino acids, such as glutamine, asparagine, serine, glycine, proline and alanine. Especially polyamino acid tracts that are encoded by repeats of a same codon are genetically unstable, due to polymerase slippage and out-of-register recombination, and thus subject to expansion and shrinkage. No less than 18 diseases are known to date to be caused by polyglutamine or polyalanine expansions. Since polyamino acid tracts are found from yeast to humans and even in bacteria the question arose whether they might also exert some positive effects, i.e., whether they might confer some kind of selective advantage. In our lab it had been shown before that a polyglutamine tract in a synthetic transcription factor can contribute to transcriptional activation, and it was postulated that expansion and shrinkage are useful to reversibly alter the activity of transcription factors in short-term evolution [Gerber HP et al, *Science* 1994, 263:808-11]. I have confirmed and extended these results by showing a clear, positive correlation between the length of a polyglutamine stretch and the transcriptional activation by the factor Gal4DBD-polyQ-VP16AD (Gal4DBD = DNA binding domain of the yeast Gal4 transcription factor, polyQ = polyglutamine stretch, VP16AD = activation domain of the viral protein VP16) in widely divergent species, namely, human cells, transgenic flies, and baker's yeast. I also wished to test the hypothesis that homopolymeric codons are genetically less stable than a mixture of codons for a homopolymeric amino acid tract, but the number of generations was probably too small to observe differences.

Project II. It is now widely recognized that a major challenge for any cell is to keep the right balance of essential trace metals and at the same time minimize the effect of non-essential ones such as cadmium, mercury, lead and silver. This is achieved by a variety of mechanisms, notably metal-specific import or export, binding to specific chaperones, storage, and detoxification by scavenging and export. Even essential metals such as copper and iron can have adverse effects if in excess, by interfering with metabolic functions via misincorporation into proteins, or by generation of reactive oxygen species due to redox cycling. Metallothioneins are small, cysteine-rich proteins which avidly bind a number of essential and non-essential heavy metals. Here I characterize metallothionein E (MtnE), the fifth and apparently ultimate member of the *Drosophila* metallothionein family. It is strongly expressed in the intestinal tract, notably in the so-called copper cells and in the iron cells of the midgut. I was also involved in the characterization of two related, small Drosophila proteins named Dumpy-30L1 and Dumpy-30L2. Dumpy-30L1 binds to, and thereby downregulates, the activity of MTF-1 (metal-responsive transcription factor-1). Accordingly, overexpression of Dumpy-30L1 rendered flies more sensitive to an excess of dietary copper or zinc. Targeted disruption of the gene for Dumpy-30L2 revealed a different phenotype in that male fertility was compromised. Furthermore, I was involved in a study led by D. Steiger on the characterization of Ctr1C, a copper importer which is strongly expressed in male gonads and contributes, together with the importer Ctr1B, to male fertility [Steiger et al, *JBC* 2010, 285(22):17089-97]. All these studies have led to a deeper understanding of various aspects of cellular metal homeostasis.

Zusammenfassung

Die vorliegende Arbeit beschäftigt sich mit zwei Themengebieten: Im ersten Teil wird die Auswirkung von Polyglutamin-Einheiten auf die Aktivität eines Transkriptionsfaktors untersucht. Während im zweiten Teil verschiedene Aspekte der Regulation von Schwermetallhomöostase und -entgiftung behandelt werden.

Projekt I: Viele Proteine, vor allem Regulatorproteine der Genexpression, enthalten Bereiche von Polyaminosäuren zum Beispiel von Glutamin, Asparagin, Serin, Glycin, Prolin und Alanin. Polyaminosäuren, welche durch Wiederholungen des selben Codons kodiert werden, sind genetisch instabil. Sich wiederholende Codons können ein Verrutschen der DNA-Polymerase oder eine fehlerhafte genetische Rekombination verursachen und führen somit zu einer Verlängerung oder einer Verkürzung der DNA-Sequenz und folglich auch der Polyaminosäure-Einheiten. Bislang sind 18 Krankheiten bekannt, welche durch eine ausufernde Verlängerung eines Polyglutamin- oder Polyalanin-Abschnitts verursacht werden. Da Polyaminosäuren von der Hefe bis zum Menschen vorkommen, stellte sich die Frage, ob Transkriptionsfaktoren, welche einen solchen Polyaminosäuren-Abschnitt enthalten, einen selektiven Vorteil haben. In unserem Labor wurde gezeigt, das ein Polyglutamin-Abschnitt in einem synthetischen Transkriptionsfaktor zur transkriptionellen Aktivität beiträgt. Es wurde vorgeschlagen, dass eine im Prinzip umkehrbare Veränderung der Polyglutaminlänge es möglich macht, die Aktivität eines Transkriptionsfaktors im Sinne einer Kurzzeit-Evolution zu modulieren [Gerber HP et al., Science 1994, 263:808-11]. In der vorliegenden Arbeit konnte ich diese Ergebnisse bestätigen und durch weitere Experimente vertiefen. In den evolutionär weit auseinander stehenden Arten, Mensch (Zellkultur), Fliege und Bäckerhefe, wird eine eindeutige positive Korrelation zwischen der Länge einer Polyglutamin-Einheit und der Aktivität des synthetischen Transkriptionsfaktors Gal4DBD-polyQ-VP16AD gezeigt (Gal4DBD: DNA-Bindedomaine des Transkriptionsfaktors Gal4 der Bäckerhefe, polyQ: Polyglutamin-Abschnitt, VP16AD: Aktivierungsdomäne des viralen Aktivatorproteins VP16). Unter anderem wollte ich auch die Hypothese testen, dass sich selbst wiederholende Codons im Vergleich zu einer Mischung von Codons, welche einen Polyaminosäure-Abschnitt kodieren, genetisch weniger stabil sind. Die Anzahl der Generationen reichte jedoch vermutlich nicht aus, um ein aussagekräftiges Ergebnis zu erhalten.

Projekt II: Jede Zelle steht vor der Herausforderung die Konzentration an essentiellen Spurenelementen auf einem bestimmten Niveau zu halten und gleichzeitig eine Schädigung durch nicht-essentielle, giftige Metalle, wie Cadmium, Quecksilber, Blei und Silber, zu vermeiden. Eine Vielzahl von zellulären Mechanismen spielen hierbei eine Rolle: metall-spezifischer zellulärer Import, Bindung an spezifische intrazelluläre Bindeproteine („Chaperone"), Speicherung und Entgiftung durch Bindung und Export. In zu hohen Konzentrationen können sogar essentielle Metalle wie Kupfer und Eisen die Zelle schädigen, indem sie unspezifisch an Proteine binden und deren Funktion beeinträchtigen und/oder Sauerstoffradikale erzeugen. Metallothioneine sind kleine, cysteinreiche Proteine, welche eine Vielzahl von Metallen mit hoher Affinität binden. In dieser Arbeit charakterisiere ich MtnE, das letzte von fünf Mitgliedern der Metallothionein-Familie in Drosophila. Dieses Protein wird im Magen/Darm-Trakt stark exprimiert, und ist vor allem in den sogenannten Kupfer- und Eisenzellen des Mitteldarmes angereichert. Zudem war ich bei einer Studie beteiligt, in der zwei nah verwandte, kleine Proteine, Dumpy-30L1 und Dumpy-30L2, in Drosophila untersucht wurden. Es wurde gezeigt, dass Dumpy-30L1 an den Transkriptionsfaktor MTF-1 („metal-responsive transcription factor-1") bindet und dadurch dessen Aktivität hemmt. Dementsprechend erhöht die Überexpression von Dumpy-30L1 in Drosophila die Sensitivität der Fliegen gegenüber Kupfer und Zink. Im Gegensatz dazu führte die Entfernung des Gens für Dumpy30L2, zu einer verminderten Fertilität männlicher Fliegen. In einer anderen Studie habe ich in Zusammenarbeit mit Dr. Dominik Steiger das Kupferimportprotein Ctr1C untersucht. Dieses Protein wird vor allem in den Keimdrüsen männlicher Fliegen exprimiert und ist neben Ctr1B, einem homologen Kupferimportprotein, für die Fertilität der Männchen notwendig [Steiger et al., JBC 2010, 285(22):17089-97]. Die Studien zu diesem Projekt haben zu einem besseren Verständnis der zellulären Metallhomöostase beigetragen.

PART I

INTRODUCTION

Tandemly repeated DNA sequences in genomes

One of the intriguing features of vertebrate genomes is the vast excess of DNA without obvious function self-perpetuating "selfish DNA" or "junk DNA", such as tandem DNA repeats, spacers and introns [Ohno, 1972; Doolittle, 1980; Orgel, 1980]. The most obvious repeated DNAs are so-called satellites, which consist of hundreds of thousands repeated sequences often located at centromeres and telomeres. Satellite DNA was first detected as a "satellite peak" in density gradient centrifugation and later identified as large tandem repeats. Shorter tandem repeats with multiple repeated sequence motifs are known as microsatellites and minisatellites. Mono, di-, tri- and tetranucleotide repeats are the main types of microsatellites, but repeats of five or six nucleotides are usually also classified as microsatellites. Tandem repeats with longer repeating units belong to minisatellites. There is also a class of mega-satellites, which includes tandem repeats of greater than 135 nucleotides [Thierry, 2008].

Tandemly repeating units of DNA repeats can be located in the non-coding intergenic regions, in the coding sequences and also in pseudogenes [Li, 2004]. Among repeats found in coding sequences, repeats with units that contain multiples of three nucleotides such as tri- and hexanucleotide repeats, are the most common [Legendre, 2007; Subramanian, 2003]. The fraction of genes containing repeats within the ORFs varies between 12 – 21% in different organisms (Table 1) [Gemayel, 2010].

Microsatellite density tends to positively correlate with genome size; among animals it is particularly high in mammals. Human genes with tandem repeats in the coding sequence are overrepresented in certain functional classes, such those for regulation of transcription, chromatin remodeling, cell-fate decision and ribosome assembly [Legendre, 2007].

Table 1. The percentage of genes with coding tandem repeats compared to total number of genes in yeast, plant, worm, fly, and human.

Species	**Genes with repeats within coding regions (%)**
Saccharomyces cerevisiae	13
Arabidopsis thaliana	14
Caenorhabditis elegans	13
Drosophila melanogaster	21
Homo sapiens	17

(Adapted from [Gemayel, 2010])

Micro- and mini-satellites are among the most variable types of DNA sequences in the genome [Hartl, 2000]. The mutation rates are as high as 10^{-3} to 10^{-7} per cell division. Most mutations represent addition or deletion of repeat units. The main sources for expansion or contraction of the repeats are DNA recombination, replication slippage (caused by secondary structure formation, such as hairpins) and DNA repair (Figure 1) [Fan, 2007; Pâques, 1998; Richard, 2000; Pearson, 2005].

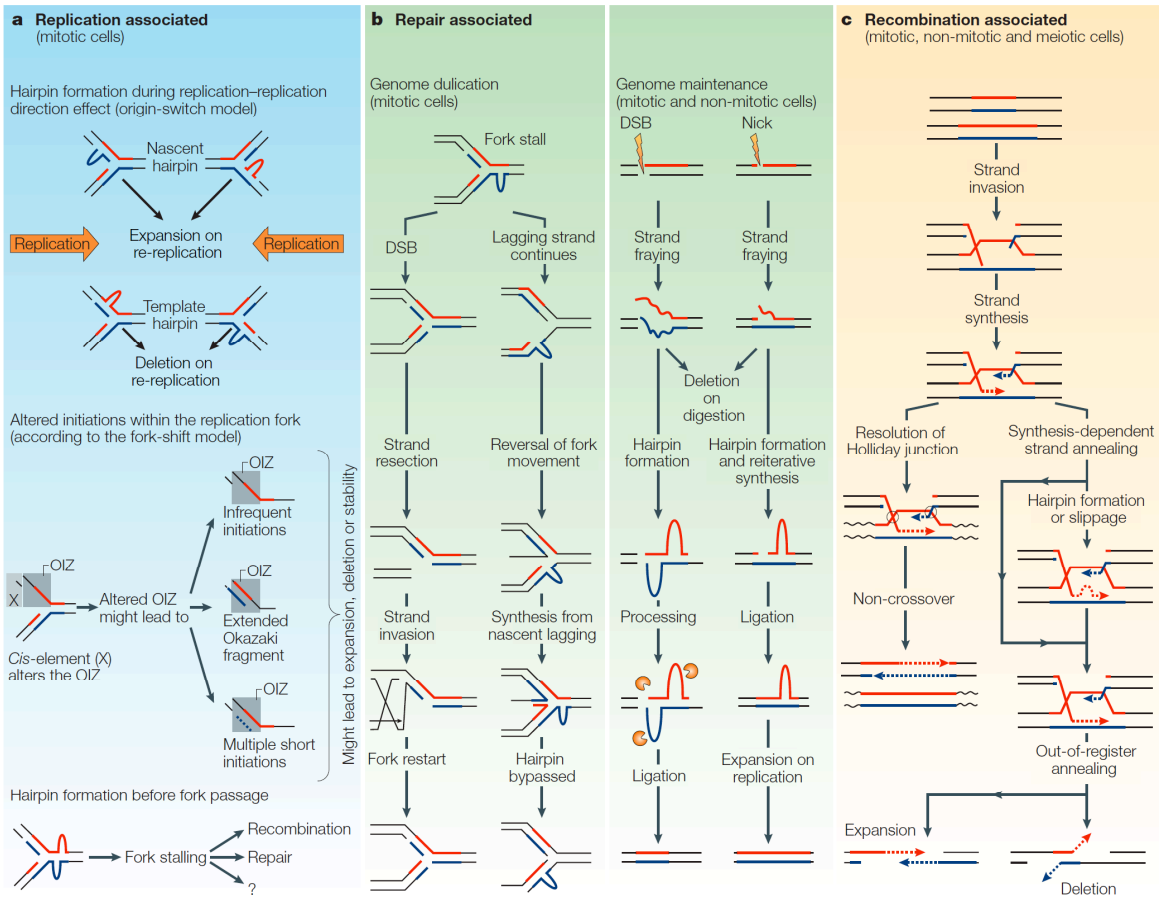


Figure 1. Scheme of molecular processes leading to DNA repeat size change.

a) The formation of unusual DNA structures and DNA slippage during lagging-strand synthesis might facilitate instability. b) Following replication fork stalling, the induction of a double-strand break (DSB) or fork reversal might result in length alterations being maintained during proceeding rounds of replication. DSB or a nick within the repeat tract might lead to strand fraying and trinucleotide-repeat-specific structures. Failure of repair to correct the alteration will result in length differences. c) Homologous recombination between allelic repeats might occur with or without the exchange of flanks. Gene conversion might occur after DSB or replication-fork blockage. Single strands from the broken repeat invade the sister tract, allowing for completion or restart of synthesis. During this period, hairpins or reiterative synthesis might occur, prompting repeat-length changes. (Adapted from [Pearson, 2005])

Unstable tandem repeats and human diseases

Nearly 30 hereditary disorders in humans result from an increase of copy number of tandem repeats in the genome (Figure 2). Fragile X syndrome (FRAXA) [Lubs, 1969; Magenis, 1970] was the first one discovered among the diseases associated with the expansion of a single trinucleotide repeat. FRAXA is caused by the expansion of CGG repeats in the 5' UTR of the *FMR* gene [Verkerk, 1991].

Repeat number changes that lead to pathogenesis can occur both in protein coding and non-coding regions of affected genes (Figure 2).

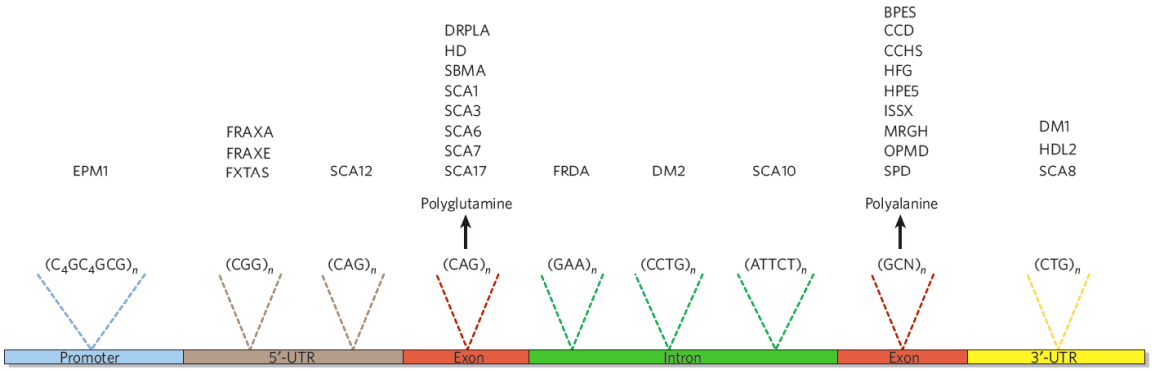


Figure 2. Gene scheme with tandem repeats responsible for human diseases.

Associated diseases are BPES, blepharophimosis, ptosis and epicanthus inversus;
CCD, cleidocranial dysplasia; CCHS, congenital central hypoventilation syndrome; DM, myotonic dystrophy; DRPLA, dentatorubral–pallidoluysian atrophy; EPM1, progressive myoclonic epilepsy 1; FRAXA, fragile X syndrome; FRAXE, fragile X mental retardation associated with FRAXE site; FRDA, Friedreich's ataxia; FXTAS, fragile X tremor and ataxia syndrome; HD, Huntington's disease; HDL2, Huntington's-disease-like 2; HFG, hand–foot–genital syndrome; HPE5, holoprosencephaly 5; ISSX, X-linked infantile spasm syndrome; MRGH, mental retardation with isolated growth hormone deficiency; OPMD, oculopharyngeal muscular dystrophy; SBMA, spinal and bulbar muscular atrophy; SCA, spinocerebellar ataxia; SPD, synpolydactyly. (Note that red sequences termed "Exon" are protein-coding, but exons normally include 5'-UTR and 3'-UTR).
(Adapted from [Mirkin, 2007])

Trinucleotide repeats tend to undergo expansions that can cause diseases, mostly neurodegenerative and neuromuscular disorders. General clinical features of triplet expansion diseases usually include a progressive loss of motor and cognitive functions, associated with neuronal cell degeneration in the central nervous system.

Within protein-coding genes, repeats of glutamine-coding CAG codons are particularly problematic. Nine neurodegenerative disorders, all of them inherited gain-of-function diseases, are caused by expansions of CAG repeats. These diseases include Huntington's disease (HD), spinobulbar muscular atrophy (SMBA or Kennedy's disease), several forms of spinocerebellar ataxia (SCA-1, -2, -3, -6, -7, -17) and dentatorubral pallidoluysian atrophy (DRPLA) (Table 2) [La Spada, 1994; Gusella, 2000; Mirkin, 2007; Orr, 2007].

HD is one of the most common inherited polyQ diseases in humans, 1 in 10'000 individuals is affected. It results from an expansion of CAG triplets, to more than the maximum 35 repeats normally present, in the first exon of the *HTT* gene (also called *HD* gene or *IT15*, "interesting transcript 15"). It encodes huntingtin, a widely expressed 348 kDa cytoplasmic protein. The consequences of the presence of aberrant huntingtin in the cells are the selective degeneration of the medium-sized striatal spiny neurons accompanied by progressive uncontrollable movements (chorea) and dementia. The presence of intranuclear and cytoplasmic aggregates of mutant huntingtin in HD brains and in animal models of the disease has been widely reported [Davies, 1998].

SBMA or Kennedy's disease is a neuromuscular disease associated with expansion of the CAG repeat in the first exon of the androgen receptor gene [La Spada, 1991; Finsterer, 2009], a key player of male sexual differentiation. The disease is characterized by progressive lower motor degeneration mostly in males in middle adult life; females are rarely affected. Symptoms include weakness, wasting, gynecomastia, progressive loss of libido and late sterility [Arbizu, 1983]. In a yeast model of Kennedy's disease it has been recently demonstrated that the dysfunction of the full-length polyQ-expanded androgen receptor is caused by its interaction with the soluble oligomers of polyQ-expanded N-terminal proteolytic fragments [Schiffer, 2008].

Spinocerebellar ataxias and DPRLA share many clinical and pathological features, such as ataxia (coordination, muscle movement dysfunction), tremor, dysarthria (motor speech disorder).

Table 2. Polyglutamine disorders caused due to the CAG expansions.

Disease	Gene product	Normal repeat length	Expanded repeat length	Main clinical features
HD	Huntingtin	6-34	36-121	Chorea, cognitive deficits, psychiatric problems
SCA1	Ataxin 1	6-44	39-82	Ataxia, slurred speech, spasticity, cognitive impairments
SCA2	Ataxin 2	15-24	32-200	Ataxia, polyneuropathy, decreased reflexes, infantile variant with retinopathy
SCA3	Ataxin 3	13-36	61-84	Ataxia, parkinsonism, spasticity
SCA6	$CACNA1_A$	4-19	10-33	Ataxia, dysarthria, nystagmus, tremors
SCA7	Ataxin7	4-35	37-306	Ataxia, blindness, cardiac failure in infantile form
SCA17	TBP	25-42	47-63	Ataxia, cognitive decline, seizures, and psychiatric problems
SBMA	Androgen receptor	9-36	38-62	Motor weakness, swallowing, gynecomastia, decreased fertility
DRPLA	Atrophin	7-34	49-88	Ataxia, seizures, choreoathetosis, dementia

(Adapted from [Orr, 2007])

All of these disorders are progressive, with a strong correlation between early disease onset and the number of triplet repeats in the specific genes.

Expression of polyQ-containing proteins in the central nervous system is sufficient to cause neurotoxicity in a mouse model [Mangiarini, 1996; Reddy, 1998], in *Drosophila* [Marsh, 2000; Warrick, 1998; Jackson, 1998], and in the nematode *Caenorhabditis elegans* [Faber, 1999; Satyal, 2000]. It was shown that the exon 1-encoded huntingtin fragment and the N-terminal 171-amino acid fragment are toxic in mouse models of HD [Schilling, 1999], whereas the N-terminal 117-aminoacid fragment, even though it contains the polyQ stretch, does not show a disease phenotype [Slow, 2005].

A common cytological feature of the triplet repeat disorders with polyQ repeat expansions is the formation of neuronal intranuclear inclusions, although cytoplasmic aggregates are also detected. Such aggregates of protein with expanded polyQs were often associated with disease progression. However, the neurons that are most vulnerable to degeneration do not necessarily correspond to the ones with inclusions [Saudou, 1998; Kuemmerle, 1999], and there is growing evidence that inclusions might have a protective role for affected cells [Arrasate, 2004; Slow, 2005; Miller, 2010]. PolyQ protein aggregation is a complex process, involving several kinds of globular or protofibrillar intermediates that might be crucial for toxicity (Figure 3) [Poirier, 2002; Sànchez, 2003].

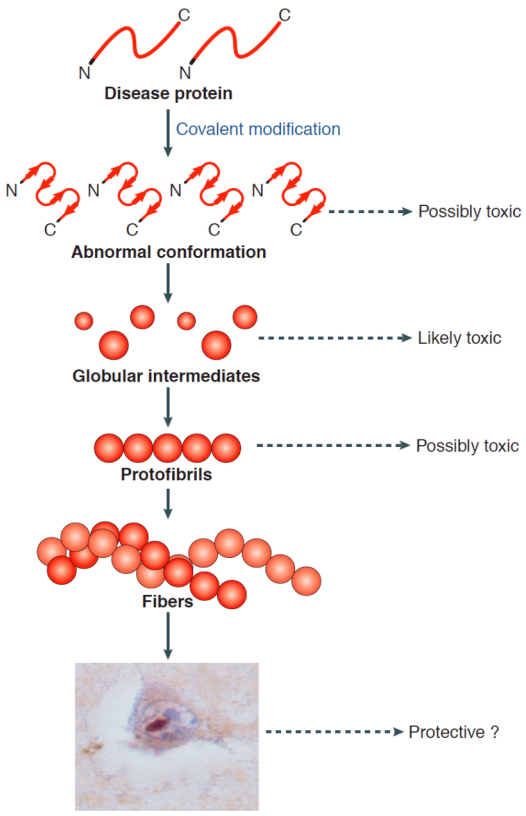


Figure 3. Hypothetical pathway of a polyQ protein aggregation.

Due to polyQ expansion the disease protein gains an abnormal conformation. Soluble, possibly toxic oligomeric (globular) intermediates may form, and from these protofibrillar structures are assembled. Amyloid fibers can then form, possibly through association of protofibrillar intermediates [Perutz, 2002], that are visible the light microscope. These aggregates are potentially cytoprotective by reducing the number of oligomeric forms.

(Adapted from [Ross, 2004])

Perutz *et al.* suggested the "polar zipper" structure for polyQ stretches: the proteins with polyQ domain form β-helices stabilized by hydrogen bonds between side chains of glutamines and the polypeptide backbone (Figure 4) [Perutz, 1994; Perutz, 2002; Merlino 2006]. An important aspect is the length of the expanded polyQ tract: a protein becomes pathogenic only after the polyQ stretch exceeds the critical size of 35-40 residues. In agreement with this notion, molecular dynamics simulations showed that only if the length of a polyQ repeat is longer than 40 glutamines, it spontaneously forms β-helices [Khare, 2005; Ogawa, 2008]. It was also found that the structure of aggregates containing an extended polyQ region is fibrillar [Scherzinger, 1999]. The β-helix structure of dimer, trimer and oligomer is more stable than that of the monomer, which is relevant for an understanding of oligomerization and aggregation [Ogawa, 2008; Barton, 2007].

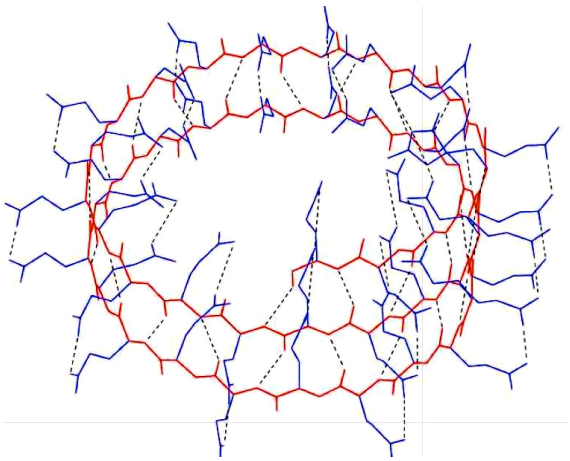


Figure 4. Computer-generated model of a poly-l-glutamine helix with 20 residues per turn.

The main chain is red and the side chains are blue. Hydrogen bonds are the black broken lines. (Adapted from [Perutz, 2002])

Transcriptional dysregulation might be a common theme in the toxicity of expanded polyQ proteins and thus, in the pathogenesis of polyQ disorders. Expanded polyQ proteins can interact with several transcriptional regulators and are able to interfere with their normal function by binding to them as monomers, or by recruiting them into aggregates [Okazawa, 2003; Helmlinger, 2006; Kazantsev, 1999]. Ectopic interactions with polyQ tracts which led to the impairment of transcription were described for CREB-binding protein (CBP) [Nucifora, 2001], p53, Sp1 and TAFII130 [Dunah, 2002; Freiman, 2002].

Accumulation of mutant polyQ proteins that form unusual cellular structures recruits quality control mechanisms, such as the ubiquitin-proteasome system, and chaperones, which might ameliorate pathogenic effects. For example, recently it was reported that HSPB7, a member of heat-shock protein subfamily HSPB, strongly suppressed polyQ aggregation and prevented polyQ-induced toxicity in mammalian cells; furthermore, overexpression of HSPB7 resulted in a rescue of eye degeneration in a *Drosophila* model of polyQ diseases [Vos, 2010]. Several pharmacological therapies involve small chemical inducers of heat shock transcription factor 1 (HSF1) such as geldanamycin and its derivative 17-AAG, which induce multiple endogenous molecular chaperones. Such inducers have been proven to be effective not only in models of polyQ diseases but also of other neurodegenerative diseases [Fujikake, 2008].

An RNA based gain-of-function mechanism might also be involved in the molecular pathogenesis in the case of SCA1, SCA3 [Li, 2008, Li, 2010].

Protein context and post-translational modifications such as phosphorylation, acetylation, sumoylation and palmitoylation, also influence the neurotoxicity of the proteins with elongated polyQ tracts [Pennuto, 2009].

Polyglutamine stretches in transcriptional activation domains

Expression of genes is regulated by transcription factors, which can function alone or in a complex with other proteins by promoting or preventing the recruitment of RNA polymerase. Transcriptional activators are typically containing a DNA-binding domain, which is responsible for the recognition of *cis*-regulatory elements in promoters, enhancers or other regulatory regions of target genes, and an activation domain (AD) which is responsible for contacting transcriptional cofactors for recruitment of the transcription apparatus [Ma, 1987; Brent, 1985; Mitchell, 1989].

Activation domains of transcriptional activator proteins are usually separate from their DNA-binding domains, and have been classified on the basis of their amino acid composition (Figure 2). “Acidic” activation domains are known to be enriched in acidic and hydrophobic amino acids, for example the carboxyl-terminal 80 amino acids (aa) of the herpes simplex virus protein VP16 [Sadowski, 1988], the C-terminal activation domain of yeast Gal4 factor [Gill, 1987] and aa 330-406 of human metal-responsive transcription factor-1 (MTF-1) [Radtke, 1995; Lindert, 2009].

Many transcription factors possess glutamine-rich (Sp1; Oct1; Snf5; TBP), proline-rich (CCAAT-box-binding transcription factor 1, CTF1) and serine/threonine-rich (aa 524-620 of human MTF-1; stage specific activator protein, SSAP; class II transactivator, CIITA) activation domains.

A minimal motif, Asp-Asp-Phe-Asp-Leu, has been proposed as a prototype module of the acidic activators [Seipel, 1994]. Some activation domains described are enriched in basic amino acids [Estruch, 1994] or are rich in isoleucine [Attardi, 1993]. Activation domains can also be classified by their functional properties, according to their ability to stimulate transcription from remote (Gal4, VP16, p65, TFE3, ITF-1, ITF-2) and/or proximal (Oct1, Oct2, Sp1) positions [Seipel, 1992]. Transcriptional activation domains have been characterized also by other criteria, such as requirement of coactivators [Tanese, 1991; Berger, 1992], activity in histone antirepression assays [Croston, 1991; Laybourn, 1991; Workman, 1991], and *in vivo* interference or 'squelching' [Tasset, 1990].

Table 2. Examples of transcriptional activation domains, based on their amino acid composition.

Acidic (VP16, MTF-1), glutamine-rich (Oct1, Oct2, Sp1), proline-rich (AP2, CTF1/NF1, p65, TFE3), serine/threonine-rich (ITF-1, ITF-2) activation domains; characteristic amino acids are indicated by bold letters.

VP16	(413-490)	APPT**D**VSLG**DE**LHL**D**G**ED**VAMAHA**D**AL**DD**F**D**L**D**MLG**D**G**D**SPGPGFTPH**D**S APYGAL**D**MA**D**F**E**F**E**QMFT**D**ALGI**DE**YGG
MTF-1	(330-406)	**D**TNHSLCLS**D**LSLLST**D**S**E**LR**E**NSSTTQGQ**D**LSTISPAIIF**E**SMFQNS**DD** TAIQ**ED**PQQTASLT**E**SFNG**D**A**E**SVS**D**
Oct1	(175-269)	DL**QQ**L**QQ**L**QQQ**NLNL**QQ**FVLVHPTTNL**Q**PA**Q**FIIS**Q**TP**Q**G**QQ**GLL**Q**A**Q**NL **Q**T**Q**LP**QQ**S**Q**ANLL**Q**S**Q**PSITLTS**Q**PATPTRTIAATPI**Q**TLP**Q**S**Q**S
Oct2	(99-161)	LAGDI**QQ**LL**Q**L**QQ**LVLVPGHHL**Q**PPA**Q**FLLP**Q**A**QQ**S**Q**PGLLPTPNLF**Q**LP **QQ**T**Q**GALLTSQPR
Sp1	(340-490)	NFTTSGSSGTNS**Q**G**Q**TP**Q**RVSGL**Q**GSDALNI**QQ**N**Q**TSGGSL**Q**AG**QQ**KEGE **Q**N**QQ**T**QQQQ**ILI**Q**P**Q**LV**Q**GG**Q**AL**Q**AL**Q**AAPLSG**Q**TFTT**Q**AIS**Q**ETL**Q**NL**Q** L**Q**AVPNSGPIIIRTPTVGPNG**Q**VSW**Q**TL**Q**LQNL**Q**V**Q**NP**Q**A**Q**TITLAPM**Q**GV
AP2	(31-76)	QLGTVGQS**P**YTSA**PP**LSHT**P**NADFQ**PP**YF**PPP**YQ**P**IY**P**QSQD**P**YSHV
CTF1/NF1	(399-499)	DLVSLACD**P**ASQQ**P**G**P**LNGSGQLKN**P**SHCLSAQNLA**PPPP**GL**P**RLAL**PP**A TK**P**ATTSZGGATS**P**TS**P**STS**PP**DTS**P**ANRSFVGLG**P**RD**P**AGIYQAQSNYLG
p65	(286-518)	FQYL**P**DTDDRHRIEEKRKRTYETFKSIMKKS**P**FSG**P**TD**P**R**PPP**RRIAV**P**S RSSASV**P**K**P**A**P**Q**P**Y**P**FTSSLSTINYDEF**P**TMVF**P**SGQISQASALA**P**A**PP**Q VL**P**QA**P**A**P**A**P**AMVSALAQA**P**A**P**V**P**VLA**P**G**PP**QAVA**PP**A**P**K**P**TQAGEGT LSEALLQLQFDDEDLGALLGNSTD**P**AVFTDLASVDNSEFQQLLNQGI**P**VA **P**HTTE**P**MLMEY**P**EAITRLVTGAQR**PP**D**P**A**P**A**P**L
TFE3	(520-570)	FGRASQALT**PPP**GKASAQ**P**L**P**A**P**EAAITTG**P**TGSA**P**NS**P**IALLTIGSSSE
ITF-1	(12-480)	TDKEL**S**DLLDF**S**MMFPLPV**T**NGKGRPA**S**LAGAQFGGSGLEDRP**SS**G**S**WG**S** GDQ**SSSS**FDPSR**T**F**S**EG**T**HF**T**E**S**H**SS**L**SSST**FLGPGLGGK**S**GERGAYA**S**F GRDAGVGGL**T**QAGFL**S**GELALN**S**PGPL**S**P**S**GMKG**TS**QYYP**S**Y**S**G**SS**RRRA ADG**S**LD**T**QPKKVRKVPPGLP**SS**VYPP**SS**GEDYGRDA**T**AYP**S**AK**T**P**SST**YP APFYVADG**S**LHP**S**AELW**S**PPGQAGFGPMLGGG**SS**PLPLPPG**S**GPVG**SS**G**S** **SST**FGGLHQHERMGYQLHGAEVNGGLP**S**A**SS**F**SS**APGA**T**YGGV**SS**HTPPV **S**GADSLLG**S**RG**TT**AG**SS**GDALGKALA**S**IY**S**PDH**SS**NNF**SSS**P**ST**PVG**S**PQ GLAG**TS**QWPRAGAPGAL**S**P**S**YDGGLHGLQ**S**KIEDHLDEAIHVLR**S**HAVG**T** AGDMH**T**LLPGHGALA**S**GF**T**GPM**S**LGGRHAGLVGG**S**HPEDGLAG**STS**LMHN HAALP**S**QPG**T**LPDL**S**RPPD
ITF-2	(2-451)	HQQRMAALG**T**DKEL**S**DLLDF**S**AMF**S**PPV**SS**GKNGP**TS**LA**S**GHF**T**G**S**NVED R**SSS**G**S**WGNGGHP**S**P**S**RNYGDG**T**PYDHM**TS**RDLG**S**HDNL**S**PPFVN**S**RIQ**S** K**T**ERG**S**Y**SS**YGRE**S**NLQGCHQQ**S**LLGGDMDMGNPG**T**L**S**P**T**KPG**S**QYYQY**S** **S**NNPRRRPLH**SS**AMEVQ**T**KKVRKVPPGLP**SS**VYAP**S**A**ST**ADYNRD**S**PGYP **SS**KPA**TST**FP**SS**FFMQDGHH**SS**DPW**SSSS**GMNQPGYAGMLGN**SS**HIPQ**SS** **S**YC**S**LHPHERL**S**YP**S**H**SS**ADIN**SS**LPPM**ST**FHR**S**G**T**NHY**STSS**CTPPANG **T**DSIMANRG**S**GAAG**SS**Q**T**GDALGKALA**S**IY**S**PDH**T**NN**S**F**SS**NP**ST**PVG**S**P P**S**L**S**AG**T**AVW**S**RNGGQA**SSS**PNYEGPLHSLQSRIEDRLERLDDAIHVLRN HAVGP**ST**AMPGGHGDMHGIIGP**S**HNGAMGGLG**S**GYGTGLL**S**ANRH**S**LMV

Simple homopolymeric repeats of glutamine or proline residues can also function as activation domains in mammalian cells when linked to the yeast Gal4 DNA-binding domain [Gerber, 1994]. Many transcription factors/cofactors contain polyQ stretches including TBP (TATA-binding protein), CBP (CREB-binding protein), Gal11 (subunit of yeast mediator complex), FoxP2 (forkhead box protein P2, involved in the developmental control of the central nervous system, linked to language and speech), N-Oct3 (nervous system-specific octamer-binding transcription factor), glucocorticoid receptor and androgen receptor.

Despite the tendency to expand and cause neurodegeneration, polyQs are conserved features of genomes from yeast to humans and triplet repeats encoding polyQs (CAG and CAA) are predominantly found in genes encoding transcription factors. There are differences in representation of triplet repeats in human genome. Some triplets, such as AGC (= CAG), CGG, and AAT are abundant, while the others (ACG, ACT) are essentially none [Künzler, 1995]. Kozlowski *et al.* studied the frequency of all trinucleotide repeats in exons and made a functional association analysis as defined by gene ontology (GO) terms. They found that polyQ is the most prevalent homopolymeric amino acid tract encoded by trinucleotide repeats. Especially striking is the association of several groups of trinucleotide repeat-containing genes with transcription-related GO terms. These groups include genes with triplet repeats, such as ACC, CAG, CGG and CCG [Kozlowksi, 2010].

It should also be mentioned that polyalanine tracts longer than 7 alanines have been shown to be more frequent in transcription regulators in the three distantly related organisms *H. sapiens*, *D. melanogaster* and *C. elegans*, suggesting a role of polyalanine tracts in the regulation of transcription [Lavoie, 2003]. To date, 9 human hereditary diseases are known to be associated with polyalanine expansions in transcriptional regulator proteins [Brown, 2004]. In contrast to polyglutamine tracts, polyalanines have been shown to act as transcriptional repressors [Lanz, 1995; Albrecht, 2005].

References

1. **Albrecht, A., and S. Mundlos.** 2005. The other trinucleotide repeat: polyalanine expansion disorders. Curr Opin Genet Dev **15:**285-293.

2. **Arbizu, T., J. Santamaría, J. Gomez, A. Quílez, and J. Serra.** 1983. A family with adult spinal and bulbar muscular atrophy, X-linked inheritance and associated testicular failure. J Neurol Sci **59:**371-82.

3. **Arrasate, M., S. Mitra, E. Schweitzer, M. Segal, and S. Finkbeiner.** 2004. Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. Nature **431:**805-10.

4. **Attardi, L., and R. Tjian.** 1993. Drosophila tissue-specific transcription factor NTF-1 contains a novel isoleucine-rich activation motif. Genes Dev **7:**1341-53.

5. **Barton, S., R. Jacak, S. Khare, F. Ding, and N. Dokholyan.** 2007. The length dependence of the polyQ-mediated protein aggregation. J Biol Chem **282:**25487-92.

6. **Berger, S., B. Piña, N. Silverman, G. Marcus, J. Agapite, J. Regier, S. Triezenberg, and L. Guarente.** 1992. Genetic isolation of ADA2: a potential transcriptional adaptor required for function of certain acidic activation domains. Cell **70:**251-65.

7. **Brent, R., and M. Ptashne.** 1985. A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. Cell **43:**729-736.

8. **Brown, L. Y., and S. A. Brown.** 2004. Alanine tracts: the expanding story of human illness and trinucleotide repeats. Trends Genet **20:**51-58.

9. **Croston, G., L. Kerrigan, L. Lira, D. Marshak, and J. Kadonaga.** 1991. Sequence-specific antirepression of histone H1-mediated inhibition of basal RNA polymerase II transcription. Science **251:**643-9.

10. **Davies, S., K. Beardsall, M. Turmaine, M. DiFiglia, N. Aronin, and G. Bates.** 1998. Are neuronal intranuclear inclusions the common neuropathology of triplet-repeat disorders with polyglutamine-repeat expansions? Lancet **351:**131-3.

11. **Doolittle, W., and C. Sapienza.** 1980. Selfish genes, the phenotype paradigm and genome evolution. Nature **284:**601-3.

12. **Dunah, A., H. Jeong, A. Griffin, Y. Kim, D. Standaert, S. Hersch, M. Mouradian, A. Young, N. Tanese, and D. Krainc.** 2002. Sp1 and TAFII130 transcriptional activity disrupted in early Huntington's disease. Science **296:**2238-43.

13. **Estruch, J., L. Crossland, and S. Goff.** 1994. Plant activating sequences: positively charged peptides are functional as transcriptional activation domains. Nucleic Acids Res **22:**3983-9.

14. **Faber, P., J. Alter, M. MacDonald, and A. Hart.** 1999. Polyglutamine-mediated dysfunction and apoptotic death of a Caenorhabditis elegans sensory neuron. Proc Natl Acad Sci U S A **96:**179-84.

15. **Fan, H., and J. Chu.** 2007. A brief review of short tandem repeat mutation. Genomics Proteomics Bioinformatics **5:**7-14.

16. **Finsterer, J.** 2009. Bulbar and spinal muscular atrophy (Kennedy's disease): a review. Eur J Neurol **16:**556-61.

17. **Freiman, R., and R. Tjian.** 2002. Neurodegeneration. A glutamine-rich trail leads to transcription factors. Science **296:**2149-2150.

18. **Fujikake, N., Y. Nagai, H. Popiel, Y. Okamoto, M. Yamaguchi, and T. Toda.** 2008. Heat shock transcription factor 1-activating compounds suppress polyglutamine-induced neurodegeneration through induction of multiple molecular chaperones. J Biol Chem **283:**26188-97.

19. **Gemayel, R., M. Vinces, M. Legendre, and K. Verstrepen.** 2010. Variable Tandem Repeats Accelerate Evolution of Coding and Regulatory Sequences. Annu Rev Genet.

20. **Gerber, H., K. Seipel, O. Georgiev, M. Höfferer, M. Hug, S. Rusconi, and W. Schaffner.** 1994. Transcriptional activation modulated by homopolymeric glutamine and proline stretches. Science **263:**808-11.

21. **Gill, G., and M. Ptashne.** 1987. Mutants of GAL4 protein altered in an activation function. Cell **51:**121-6.

22. **Gusella, J. F., and M. E. MacDonald.** 2000. Molecular genetics: unmasking polyglutamine triggers in neurodegenerative disease. Nat Rev Neurosci **1:**109-115.

23. **Hartl, D.** 2000. Molecular melodies in high and low C. Nat Rev Genet **1:**145-9.

24. **Helmlinger, D., L. Tora, and D. Devys.** 2006. Transcriptional alterations and chromatin remodeling in polyglutamine diseases. Trends Genet **22:**562-70.

25. **Jackson, G., I. Salecker, X. Dong, X. Yao, N. Arnheim, P. Faber, M. MacDonald, and S. Zipursky.** 1998. Polyglutamine-expanded human huntingtin transgenes induce degeneration of Drosophila photoreceptor neurons. Neuron **21:**633-42.

26. **Kazantsev, A., E. Preisinger, A. Dranovsky, D. Goldgaber, and D. Housman.** 1999. Insoluble detergent-resistant aggregates form between pathological and nonpathological lengths of polyglutamine in mammalian cells. Proc Natl Acad Sci U S A **96:**11404-9.

27. **Khare, S., F. Ding, K. Gwanmesia, and N. Dokholyan.** 2005. Molecular origin of polyglutamine aggregation in neurodegenerative diseases. PLoS Comput Biol **1:**230-5.

28. **Kozlowski, P., M. de Mezer, and W. Krzyzosiak.** 2010. Trinucleotide repeats in human genome and exome. Nucleic Acids Res **38:**4027-39.

29. **Kuemmerle, S., C. Gutekunst, A. Klein, X. Li, S. Li, M. Beal, S. Hersch, and R. Ferrante.** 1999. Huntington aggregates may not predict neuronal death in Huntington's disease. Ann Neurol **46:**842-9.

30. **Künzler, P., K. Matsuo, and W. Schaffner.** 1995. Pathological, physiological, and evolutionary aspects of short unstable DNA repeats in the human genome. Biol Chem Hoppe Seyler **376:**201-211.

31. **La Spada, A., E. Wilson, D. Lubahn, A. Harding, and K. Fischbeck.** 1991. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. Nature **352:**77-9.

32. **La Spada, A., H. Paulson, and K. Fischbeck.** 1994. Trinucleotide repeat expansion in neurological disease. Ann Neurol **36:**814-822.

33. **Lanz, R. B., S. Wieland, M. Hug, and S. Rusconi.** 1995. A transcriptional repressor obtained by alternative translation of a trinucleotide repeat. Nucleic Acids Res **23:**138-145.

34. **Lavoie, H., F. Debeane, Q. D. Trinh, J. F. Turcotte, L. P. Corbeil-Girard, M. J. Dicaire, A. Saint-Denis, M. Pagé, G. A. Rouleau, and B. Brais.** 2003. Polymorphism, shared functions and convergent evolution of genes with sequences coding for polyalanine domains. Hum Mol Genet **12:**2967-2979.

35. **Laybourn, P., and J. Kadonaga.** 1991. Role of nucleosomal cores and histone H1 in regulation of transcription by RNA polymerase II. Science **254:**238-45.

36. **Legendre, M., N. Pochet, T. Pak, and K. Verstrepen.** 2007. Sequence-based estimation of minisatellite and microsatellite repeat variability. Genome Res **17:**1787-96.

37. **Li, L., and N. Bonini.** 2010. Roles of trinucleotide-repeat RNA in neurological disease and degeneration. Trends Neurosci **33:**292-8.

38. **Li, L., Z. Yu, X. Teng, and N. Bonini.** 2008. RNA toxicity is a component of ataxin-3 degeneration in Drosophila. Nature **453:**1107-11.

39. **Li, Y., A. Korol, T. Fahima, and E. Nevo.** 2004. Microsatellites within genes: structure, function, and evolution. Mol Biol Evol **21:**991-1007.

40. **Lindert, U., M. Cramer, M. Meuli, O. Georgiev, and W. Schaffner.** 2009. Metal-responsive transcription factor 1 (MTF-1) activity is regulated by a nonconventional nuclear localization signal and a metal-responsive transactivation domain. Mol Cell Biol **29:**6283-93.

41. **Lubs, H.** 1969. A marker X chromosome. Am J Hum Genet **21:**231-44.

42. **Ma, J., and M. Ptashne.** 1987. A new class of yeast transcriptional activators. Cell **51:**113-9.

43. **Mangiarini, L., K. Sathasivam, M. Seller, B. Cozens, A. Harper, C. Hetherington, M. Lawton, Y. Trottier, H. Lehrach, S. W. Davies, and G. P. Bates.** 1996. Exon 1 of the HD gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. Cell **87:**493-506.

44. **Magenis, R., F. Hecht, and E. Lovrien.** 1970. Heritable fragile site on chromosome 16: probable localization of haptoglobin locus in man. Science **170:**85-7.

45. **Marsh, J., H. Walker, H. Theisen, Y. Zhu, T. Fielder, J. Purcell, and L. Thompson.** 2000. Expanded polyglutamine peptides alone are intrinsically cytotoxic and cause neurodegeneration in Drosophila. Hum Mol Genet **9:**13-25.

46. **Merlino, A., L. Esposito, and L. Vitagliano.** 2006. Polyglutamine repeats and beta-helix structure: molecular dynamics study. Proteins **63:**918-27.

47. **Miller, J., M. Arrasate, B. Shaby, S. Mitra, E. Masliah, and S. Finkbeiner.** 2010. Quantitative relationships between huntingtin levels, polyglutamine length, inclusion body formation, and neuronal death provide novel insight into huntington's disease molecular pathogenesis. J Neurosci **30:**10541-50.

48. **Mirkin, S.** 2007. Expandable DNA repeats and human disease. Nature **447:**932-40.

49. **Mitchell, P., and R. Tjian.** 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science **245:**371-8.

50. **Nucifora, F. J., M. Sasaki, M. Peters, H. Huang, J. Cooper, M. Yamada, H. Takahashi, S. Tsuji, J. Troncoso, V. Dawson, T. Dawson, and C. Ross.** 2001. Interference by huntingtin and atrophin-1 with cbp-mediated transcription leading to cellular toxicity. Science **291:**2423-8.

51. **Ogawa, H., M. Nakano, H. Watanabe, E. Starikov, S. Rothstein, and S. Tanaka.** 2008. Molecular dynamics simulation study on the structural stabilities of polyglutamine peptides. Comput Biol Chem **32:**102-10.

52. **Ohno, S.** 1972. So much "junk" DNA in our genome. Brookhaven Symp Biol **23:**366-70.

53. **Okazawa, H.** 2003. Polyglutamine diseases: a transcription disorder? Cell Mol Life Sci **60:**1427-39.

54. **Orgel, L., and F. Crick.** 1980. Selfish DNA: the ultimate parasite. Nature **284:**604-7.

55. **Orr, H., and H. Zoghbi.** 2007. Trinucleotide repeat disorders. Annu Rev Neurosci **30:**575-621.

56. **Pearson, C., K. Nichol Edamura, and J. Cleary.** 2005. Repeat instability: mechanisms of dynamic mutations. Nat Rev Genet **6:**729-42.

57. **Pennuto, M., I. Palazzolo, and A. Poletti.** 2009. Post-translational modifications of expanded polyglutamine proteins: impact on neurotoxicity. Hum Mol Genet **18:**R40-7.

58. **Perutz, M., J. Finch, J. Berriman, and A. Lesk.** 2002. Amyloid fibers are water-filled nanotubes. Proc Natl Acad Sci U S A **99:**5591-5.

59. **Perutz, M., T. Johnson, M. Suzuki, and J. Finch.** 1994. Glutamine repeats as polar zippers: their possible role in inherited neurodegenerative diseases. Proc Natl Acad Sci U S A **91:**5355-8.

60. **Poirier, M., H. Li, J. Macosko, S. Cai, M. Amzel, and C. Ross.** 2002. Huntingtin spheroids and protofibrils as precursors in polyglutamine fibrilization. J Biol Chem **277:**41032-7.

61. **Pâques, F., W. Leung, and J. Haber.** 1998. Expansions and contractions in a tandem repeat induced by double-strand break repair. Mol Cell Biol **18:**2045-54.

62. **Radtke, F., O. Georgiev, H. Müller, E. Brugnera, and W. Schaffner.** 1995. Functional domains of the heavy metal-responsive transcription regulator MTF-1. Nucleic Acids Res **23:**2277-2286.

63. **Reddy, P. H., M. Williams, V. Charles, L. Garrett, L. Pike-Buchanan, W. O. Whetsell, G. Miller, and D. A. Tagle.** 1998. Behavioural abnormalities and selective neuronal loss in HD transgenic mice expressing mutated full-length HD cDNA. Nat Genet **20:**198-202.

64. **Richard, G., and F. Pâques.** 2000. Mini- and microsatellite expansions: the recombination connection. EMBO Rep **1:**122-6.

65. **Ross, C., and M. Poirier.** 2004. Protein aggregation and neurodegenerative disease. Nat Med **10 Suppl:**S10-7.

66. **Sadowski, I., J. Ma, S. Triezenberg, and M. Ptashne.** 1988. GAL4-VP16 is an unusually potent transcriptional activator. Nature **335:**563-4.

67. **Satyal, S., E. Schmidt, K. Kitagawa, N. Sondheimer, S. Lindquist, J. Kramer, and R. Morimoto.** 2000. Polyglutamine aggregates alter protein folding homeostasis in Caenorhabditis elegans. Proc Natl Acad Sci U S A **97:**5750-5.

68. **Saudou, F., S. Finkbeiner, D. Devys, and M. Greenberg.** 1998. Huntingtin acts in the nucleus to induce apoptosis but death does not correlate with the formation of intranuclear inclusions. Cell **95:**55-66.

69. **Scherzinger, E., A. Sittler, K. Schweiger, V. Heiser, R. Lurz, R. Hasenbank, G. Bates, H. Lehrach, and E. Wanker.** 1999. Self-assembly of polyglutamine-containing huntingtin fragments into amyloid-like fibrils: implications for Huntington's disease pathology. Proc Natl Acad Sci U S A **96:**4604-9.

70. **Schiffer, N., J. Céraline, F. Hartl, and S. Broadley.** 2008. N-terminal polyglutamine-containing fragments inhibit androgen receptor transactivation function. Biol Chem **389:**1455-66.

71. **Schilling, G., M. Becher, A. Sharp, H. Jinnah, K. Duan, J. Kotzuk, H. Slunt, T. Ratovitski, J. Cooper, N. Jenkins, N. Copeland, D. Price, C. Ross, and D. Borchelt.** 1999. Intranuclear inclusions and neuritic aggregates in transgenic mice expressing a mutant N-terminal fragment of huntingtin. Hum Mol Genet **8:**397-407.

72. **Seipel, K., O. Georgiev, and W. Schaffner.** 1994. A minimal transcription activation domain consisting of a specific array of aspartic acid and leucine residues. Biol Chem Hoppe Seyler **375:**463-70.

73. **Seipel, K., O. Georgiev, and W. Schaffner.** 1992. Different activation domains stimulate transcription from remote ('enhancer') and proximal ('promoter') positions. EMBO J **11:**4961-8.

74. **Slow, E., R. Graham, A. Osmand, R. Devon, G. Lu, Y. Deng, J. Pearson, K. Vaid, N. Bissada, R. Wetzel, B. Leavitt, and M. Hayden.** 2005. Absence of behavioral abnormalities and neurodegeneration in vivo despite widespread neuronal huntingtin inclusions. Proc Natl Acad Sci U S A **102:**11402-7.

75. **Subramanian, S., V. Madgula, R. George, R. Mishra, M. Pandit, C. Kumar, and L. Singh.** 2003. Triplet repeats in human genome: distribution and their association with genes and other genomic regions. Bioinformatics **19:**549-52.

76. **Sánchez, I., C. Mahlke, and J. Yuan.** 2003. Pivotal role of oligomerization in expanded polyglutamine neurodegenerative disorders. Nature **421:**373-9.

77. **Tanese, N., B. Pugh, and R. Tjian.** 1991. Coactivators for a proline-rich activator purified from the multisubunit human TFIID complex. Genes Dev **5:**2212-24.

78. **Tasset, D., L. Tora, C. Fromental, E. Scheer, and P. Chambon.** 1990. Distinct classes of transcriptional activating domains function by different mechanisms. Cell **62:**1177-87

79. **Thierry, A., C. Bouchier, B. Dujon, and G. Richard.** 2008. Megasatellites: a peculiar class of giant minisatellites in genes involved in cell adhesion and pathogenicity in Candida glabrata. Nucleic Acids Res **36:**5970-82.

80. **Verkerk, A., M. Pieretti, J. Sutcliffe, Y. Fu, D. Kuhl, A. Pizzuti, O. Reiner, S. Richards, M. Victoria, and F. Zhang.** 1991. Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell **65:**905-14.

81. **Vos, M., M. Zijlstra, B. Kanon, M. van Waarde-Verhagen, E. Brunt, H. Oosterveld-Hut, S. Carra, O. Sibon, and H. Kampinga.** 2010. HSPB7 is the most potent polyQ aggregation suppressor within the HSPB family of molecular chaperones. Hum Mol Genet.

82. **Warrick, J., H. Paulson, G. Gray-Board, Q. Bui, K. Fischbeck, R. Pittman, and N. Bonini.** 1998. Expanded polyglutamine protein forms nuclear inclusions and causes neural degeneration in Drosophila. Cell **93:**939-49.

83. **Workman, J., I. Taylor, and R. Kingston.** 1991. Activation domains of stably bound GAL4 derivatives alleviate repression of promoters by nucleosomes. Cell **64:**533-44.

Title:

Polyglutamine stretches as universal transcriptional activation domains

Authors: Lilit Atanesyan[1, 2], Oleg Georgiev[1], Walter Schaffner[1]

[1] Institute of Molecular Life Sciences, University of Zurich,
Winterthurerstrasse 190, CH-8057, Zurich, Switzerland

[2] Life Science Zurich Graduate School, University of Zurich/ETH Zurich, Switzerland

To whom correspondence should be addressed:

Walter Schaffner
Institute of Molecular Life Sciences, University of Zurich,
Winterthurerstrasse 190, CH-8057 Zurich, Switzerland;
phone: +41 635 31 50/51; fax: +41 44 635 68 11; e-mail: walter.schaffner@imls.uzh.ch

Abstract

Poly-amino acid repeats, especially long stretches of glutamines (polyQs), are common features of transcription factors, co-factors and cell-signaling proteins and are prone to expansion and shrinkage. Especially unstable are polyQs encoded by CAG triplets. Nine neurodegenerative diseases are known to be caused by expansions of CAG repeats. Transcriptional dysregulation is a likely mechanism of pathogenesis due to the ability of proteins with expanded polyQs to interact with several transcription factors and interfere with their normal function. Simple homopolymeric stretches of prolines and glutamines can themselves activate transcription in mammalian cells. We used a model transcription factor composed of the DNA-binding domain of Gal4 and the activation domain of VP16 protein, which are connected via a polyQ stretch of various length (Gal4DBD-polyQ-VP16AD). When these transcription factors were studied in mammalian and *Drosophila* cell culture, in transgenic flies (*Drosophila melanogaster*) and in budding yeast (*Saccharomyces cerevisiae*), polyQs contributed to transcriptional activation in a length-dependent manner. Finally, using a two-hybrid transcription assay in mammalian cells we demonstrate a protein-protein interaction mediated by polyQ domains.

Introduction

Homopolymeric glutamine stretches (polyQs) are the most prevalent homo-amino acid (homo-aa) tracts encoded by trinucleotide repeats, accounting for 19% of all homo-aa tracts in the human proteome. PolyQs are also the most frequent type of homo-aa tracts in the *D. melanogaster* proteome. At least thirty *D. melanogaster* genes conserved in all twelve *Drosophila* species possess conserved polyQ tracts of twenty residues or longer [Bettencourt, 2007]. Functional association analysis showed that the most prevalent tandem nucleotide repeats localized in ORFs (CGG coding for alanine, CAG - glutamine, CCG - proline, AGG - glutamic acid, ACC - histidine) are predominantly associated with transcription-related functions [Kozlowski, 2010]. The activity of model transcription factors can be increased by short homopolymeric stretches of glutamine or proline [Gerber, 1994]. Examples of polyQ-containing proteins, which play important roles in the transcription of a variety of genes in yeast, flies and humans are listed in Table 1.

Especially when polyQ tracts are encoded by CAG triplets they tend to be unstable. The appearance of expanded polyQ stretches in proteins is a hallmark of disease progression, such as Huntington's disease, spinal bulbar muscular atrophy (Kennedy's disease), spinocerebellar ataxia-1, -2, -3, -6, -7, -17, dentatorubral pallidoluysian atrophy [La Spada, 1994; Orr, 2007; Gusella, 2000; Mirkin, 2007]. Usually the critical length of a polyQ stretch is 35-40 residues. Above this threshold polyQ tracts tend to self-aggregate into fibrils, wherein the polyQs form stable "polar zippers" of β-barrels, suggesting a so-called β-helix nanotube model [Perutz, 1994; Perutz 2002; Merlino, 2006]. Intraneural aggregates or inclusions assembled of proteins with expanded polyQ stretch are the main feature of polyQ diseases. These insoluble aggregates might have a protective role in affected cells by preventing many small, more deleterious oligomeric aggregates [Scherzinger, 1999; Arrasate, 2004; Miller, 2010; Poirier, 2002; Sànchez, 2003].

Transcriptional alterations are known to be implicated in several polyQ diseases [Riley, 2006; Sugars, 2003]. Expanded polyQ proteins can interact with polyQ-containing or non-containing transcription factors/co-factors and interfere with their normal function mostly by sequestration of a target protein by polyQ protein monomers, or recruitment into aggregates [Okazawa, 2003; Helmlinger, 2006; Kazantsev, 1999]. Aberrant polyQ

interactions have been described for CREB-binding protein (CBP), p53, Sp1, TAFII130 [Nucifora, 2001; Dunah, 2002; Freiman, 2002], which are associated with impaired transcription.

Many regulators of transcription contain glutamine-rich activation domains that are important in protein-protein interactions and gene expression regulation [Tanese, 1993]. Despite the highly "sticky" nature of expanded polyQ-containing proteins, in a yeast two-hybrid study no interaction was observed between normal-length (30Q) polyQ tracts [Oma, 2007]. The size of the polyQ might be considered as an important feature of polyQ-containing proteins for certain conformations and interactions. The polyQ size, due to its instability, might also be an "evolutionary tool" or a "tuning knob" for the modulation of transcription factor activity [Gerber, 1994].

Materials and methods

Construction of plasmids

The reporter and reference plasmids in all *in vivo* mammalian and *Drosophila* cell culture transcription experiments are based on the rabbit beta-globin OVEC (oligonucleotide vector) system [Westin, 1987]. Construction of expression vectors containing aa 1-93 of DNA-binding domain of Gal4 (Gal4DBD) and the aa 413-493 of VP16 ($VP16AD^{80}$), flanked by poly[CAG] or by poly[AGC] stretches of various lengths indicated in the Figure 1, was performed by inserting the 183 bp metallothionein B promoter (pMtnB) fragment (SalI, SacI) from the dMtoOVEC plasmid [Zhang, 2001] into pAc-attB-Gal4DBD-nQ-VP16AD plasmid (nQ is the polyQ stretch; n indicates the number of glutamine residues), thus replacing the actin promoter. The pActin-attB-Gal4-nQ-VP16AD plasmid was made by inserting the Gal4DBD-nQ-VP16AD coding sequence into pAttB vector [Bischof, 2007], followed by insertion of the actin promoter. We could not obtain transgenic *Drosophila*, expressing the model transcription factor from actin promoter, possibly due to the toxicity of the construct. Therefore the inducible metallothionein B promoter was used. The construction of pCMV-Gal4DBD-nQ-VP16AD plasmids was described previously [Gerber, 1994]. "Prey" and "bait" constructs for two-hybrid studies (pCMV-Gal4DBD-nQ and pCMV-nQ-VP16) were generated by removing either the Gal4DBD or the VP16AD moieties from pCMV-Gal4DBD-nQ-VP16 constructs.

For transcription experiments in *S. cerevisiae* we used the high copy number pAS2ΔΔ vector, where fusions of Gal4DBD with the N-terminal 40 aa of VP16AD ($VP16AD^{40}$) were driven by the ADH1 promoter. $VP16AD^{40}$ was generated by introducing several stop codons into $VP16AD^{80}$.

S1 nuclease protection assay

Cells of an exponentially growing 10 cm culture dish were transfected with 0.5-6 μg of the particular expression clones, 10 μg of reporter and 1-5 μg of reference plasmid, as indicated in the figure legends. Herring sperm DNA was added as a carrier DNA to each sample to a total of 20 μg. The isolation of RNA and the S1 nuclease protection assay were done as described previously [Weaver, 1979; Westin, 1987], typically with 100 - 150 μg of total RNA per sample. Signals were visualized using the fluorescent image analyzer FLA-7000 and quantified using the ImageGauge software (Fujifilm Life Science). Reporter signals were normalized to the reference signals.

Cell culture and transfection

Culturing and transfection of HEK293 and HEK293T cells were carried out in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (Biochrom AG), 100 U/ml penicillin, 100 μg/ml streptomycin and 2 mM L-glutamine (Invitrogen). Reporter and reference genes were transfected into cells together with the expression vectors by the calcium-phosphate co-precipitation method [Graham, 1973]. After 16 hours cells were washed with tris-buffered saline (TBS) and incubated for 24 hours in growth medium before harvesting them for RNA extraction or nuclear protein extract preparation. For transfections, a total of 20 μg DNA was used per 10 cm cell culture dish. *Drosophila* Kc cells were cultured in Schneider's *Drosophila* medium which contains 10% heat-inactivated fetal bovine serum (Biochrom AG), 100 U/ml penicillin, 100 μg/ml streptomycin and 2 mM L-glutamine (Invitrogen). As for mammalian cells, the calcium-phosphate co-precipitation method was used for transfection, and 30 μg DNA was transfected per 10 cm dish. The cells were harvested for further manipulations 48 or 72 hours after transfection.

Liquid lacZ assay for *Drosophila* larvae

For measuring beta-galactosidase activity, samples of three wandering larvae were transferred to a 1.5 ml reaction tube and lysed by successive freeze-thaw cycles in liquid nitrogen. 100 μl of Z-buffer (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM

$MgSO_4$) were added to each tube and vortexed for 20 seconds. 700 µl of Z-buffer (containing 50 mM β-mercaptoethanol) were then added to each sample. 160 µl of 4 mg/ml ONPG (ortho-nitrophenyl-β-galactoside) substrate were added to start the reaction. After 5 min at 30°C 400 µl of 1 M Na_2CO_3 were added and the samples centrifuged for 10 min at 13'000 rpm. The OD of the clear supernatant was measured at 420 nm.

Yeast strains and gene expression

Growth and handling of yeast was carried out with standard techniques. Yeast was grown in a rich YPD medium (2% glucose, 2% bacto-tryptone, 1% yeast extract) or in a synthetic drop-out medium when selection for plasmid-dependent growth was required (2% glucose, 0.17% yeast nitrogen base, 0.5% ammonium sulfate, 1 x aa). The following *S. cerevisiae* strains were used: Y190 (MATa, *gal4-542*, *gal80-538*, *his3*, *trp1-901*, *ade2-101*, *ura3-52*, *leu2-3,112*, URA3::GAL1-LacZ, Lys2::GAL1-HIS3cyhr) and Y187 (*MAT*α, *ura3-52*, *his3-200*, *ade2-101*, *trp1-901*, *leu2-3, 112*, *gal4Δ*, *met–*, *gal80Δ*, *MEL1*, *URA3*::$GAL1_{UAS}$ -$GAL1_{TATA}$-*lacZ)*. Gene expression in yeast was monitored by an *in vivo* plate assayusing X-gal (bromo-chloro-indolyl-galactopyranoside) in the medium and by a quantitative liquid culture assay using ONPG as substrate (CLONTECH protocols).

Fly food, transgenic fly strains

1 liter of fly food was composed of 55 g cornmeal (Maisgriess 54.401.025, Meyerhans Hotz AG), 100 g yeast (Hefe Schweiz AG), 75 g sugar (Dextrose monohydrate), 8 g agar (Insectagar type ZN5) and 15 ml anti-fungal Nipagin (nipagin 33 g/l, nipasol 66 g/l in 96% ethanol). For the induction of the MtnB promoter, food was supplemented with 100 µM $CuSO_4$. Transgenes encoding chimeric transcription factors with or without a polyQ stretch were introduced into *Drosophila* by the ϕC31 integrase-mediated transgenesis technique [Bischof, 2007; Fish, 2007]. All the transgenes were inserted at the *86Fb* locus on the 3rd chromosome. As reporter fly strains *UAS-GFP* (transgene on the 2nd chromosome) and *UAS-lacZ* (transgene on the 3rd chromosome) were used.

Results

PolyQs contribute to transcriptional activation in a reporter system driven by a synthetic transcription factor

A chimeric transcription factor composed of aa 1-93 of the DNA-binding domain of yeast Gal4 transcription factor (Gal4DBD) and the carboxyl-terminal 80 aa of the herpes simplex virus protein VP16 ($VP16AD^{80}$), which are linked via a polyQ stretch: Gal4DBD-polyQ-$VP16AD^{80}$ (Figure 1) was studied. The lengths of polyQ stretches were chosen such that they correspond to naturally occurring non-pathogenic size (20Q), close to pathogenic size (40Q) and expanded abnormal length (86Q). In *Drosophila* cell culture, the polyQ tract had a positive effect on transcription efficiency. Levels of reporter gene transcripts were highest with 40Q and also elevated with 20Q- and 86Q-containing effector proteins compared to the one without the polyQ stretch (Figure 2A).

By the ϕC31 integrase-mediated transgenesis technique [Bischof, 2007; Fish, 2007] we introduced a chimeric transcription factor with or without a polyQ stretch (Figure 1) into *Drosophila*. A GFP reporter system was used to study the activity of Gal4DBD-polyQ-$VP16AD^{80}$ synthetic transcription factor. We found that polyQs also contributed to transcriptional activation in the fruitfly, whereby flies with the 86Q-containing transcription factor show a stronger GFP signal than the factor with 20Q (Figure 2B). A positive correlation between transactivation and the length of the polyQ tract was also observed with an UAS-lacZ reporter (Figure 2C).

Screening of the *S. cerevisiae* genome database revealed that polyQ tracts are the most common homopolymeric amino acid stretches among the 5885 annotated proteins (Table 2). Together with poly-asparagine, poly-proline and poly-glutamic acid, polyQs are often found in proteins that are involved in transcription and regulation of transcription. Therefore, we decided to also test polyQs in yeast. An *in vivo* plate assay using X-Gal in the medium showed that in *S. cerevisiae* the synthetic transcription factors carrying polyQ tracts of various sizes activate transcription of a *LacZ* reporter gene in a polyQ length-dependent manner (Figure 2D). Only a 40 aa subsegment of the VP16 activation domain ($VP16^{40}$) was used in our yeast studies to avoid selection against the very potent

and possibly toxic full-length activation domain [Sadowski, 1988], and also to retain the possibility to increase transactivation by introducing a polyQ stretch.

Polyglutamine but not polyserine stimulates transcription

Poly[CAG] can be translated into three different homopolymeric amino acid stretches, depending on the reading frame: polyglutamine, polyserine and polyalanine. Homopolymeric serine tracts (polyS) have been reported to behave in a neutral manner with regard to transcriptional activity, whereas polyalanines act as transcriptional repressors [Lanz, 1995]. In line with this notion, tracts of 20 or 86 serines between Gal4DBD and $VP16AD^{80}$ did not increase transcriptional activity of the chimeric transcription factor (Figure 3A). The polyserine tract also acts in a neutral manner in *Drosophila* (Figure 3B). These findings show that stimulation of transcription by polyQ tracts located between DNA binding domain and activation domain is not merely due to a linker effect.

Transactivation via polyQ interactions tested in a mammalian two-hybrid system

In order to test for polyQ-polyQ interactions *in vivo*, we generated a two-hybrid system in mammalian cell culture. In this system, polyQ fused to the Gal4DBD acts as a "bait" and polyQ fused to the $VP16AD^{80}$ acts as a "prey" (Figure 4). PolyQ-polyQ interaction can induce transcription in this system, as shown by the S1-nuclease protection assay, whereby longer polyQs on one or both partners are better inducers of transcription (Figure 5). Electrophoretic mobility shift assays (EMSA) show that the mRNAs of different "bait" constructs are translated, but the interaction between polyQ-containing bait and prey proteins is probably weak, because no supershift with the "prey" is observed (Figure 6).

Discussion

Despite their genetic instability and their tendency to expand and cause neurodegeneration, polyQs are conserved features of genomes from yeast to humans and are predominantly found in genes encoding transcription factors and transcriptional regulators (enrichment in proteins of this function for human genes has an adjusted *P*-value of 8.05×10^{-9} [Legendre, 2007]). The presence of polyQ tracts in DNA binding transcription factors and co-factors suggests a functional importance of polyQs. It was previously shown that in mammalian cells glutamine-rich domains synergize with acidic activation domains [Seipel, 1992]. Indeed, by introduction of a polyQ stretch into a synthetic transcription factor increased its transcriptional activity. By varying the size of the polyQ stretch within a transcription factor we could modulate its activity. Our studies reveal that polyQ domains can be regarded as universal activation domains because they worked in all our experimental models: mammalian cells, *Drosophila* and yeast.

Transcriptional activation *in vivo* is optimal for a certain number of repeats; in our case 40Qs resulted in the highest transcriptional activity in mammalian cells and in *Drosophila* (Figure 2A, 3B). 86Qs caused a stronger activation than 20Qs in all systems studied, but still less than 40Qs. This might be explained by the propensity of oversized polyQ stretches for protein aggregation which would make part of the transcription factor insoluble and unable to participate in the process of gene expression. In other words, a fraction of the transcription factor with expanded polyQs would form aggregates, at which would be sequestered and thus not engaged in transcription. In yeast cells, the synthetic transcription factor with 86Qs is still the most potent one (Figure 2D). This might indicate that in different cellular environments the threshold for aggregation varies. Chaperones, such as heat shock proteins (HSPs) that regulate protein folding [Schaffar, 2004; Neef, 2010; Robertson, 2010], might function differently. We note that unlike mammals and *Drosophila*, yeast contains a very aggressive chaperone termed hsp104 that is particularly suited to dissolve protein aggregates [Krobitsch, 2000].

Another possible explanation for this difference might be the requirement and availability of polyQ-interacting components of the general transcription machinery, such as TFIID. It is known that expanded polyglutamine stretches interact with metazoan

transcription factors such Sp1, TAFII130 (TAFII110 in *Drosophila*), and p300/CREB, thus interfering with CREB-dependent transcription [Shimohata, 2000; Dunah, 2002].

In our two-hybrid experiments we saw positive effects in human HEK293T cells but neither in some other mammalian cells, nor in *Drosophila* cells and transgenic *Drosophila* strains (data not shown). The result that polyQ interactions induce transcriptional activation in HEK293T cells might be due to the propensity of these cells to take up transfected DNA very efficiently and then produce high amounts of proteins, thus making their interaction more likely to happen. We do not exclude that proteins other than polyQ-containing prey and bait, such as p300 or TAFII130, can also be involved in this interaction, because there is a weak transactivation with only the polyQ-carrying prey protein (data not shown).

Triplet usage (CAA or CAG) is particularly important for genetic stability of repeat size and therefore in polyQ-associated neurological disorders. Pure CAG repeats as found in the polyQ tracts of TBP (TATA-binding protein) in the SCA17 locus (spinocerebellar ataxia type 17) are less stable than repeats encoded by a mix of interspersed CAG and CAA triplets [Gao, 2008]. In a transcription factor needed for speech and language development in humans, such as FoxP2 (Forkhead box protein P2), a long polyQ stretch (35-41 glutamines) seems very stable in vertebrates [Bruce, 2002; Webb, 2005], probably because of mixed CAG/CAA triplets.

Finally, the question arises: what is the role of the polyQ size variation in natural transcription factors or other proteins? Can poly amino acid repeats, especially CAG-encoded glutamine repeats, be exploited to modulate the transcription factor activity under selective pressure? It remains to be experimentally shown that transcription factors and regulatory proteins will adapt to certain environmental conditions by expanding or shrinking the polyQ stretch size.

Acknowledgements

We are grateful to Dr. Johannes Bischof for kindly providing many fly strains and to Prof. Bernhard Dichtl for yeast strains and valuable guidance. We also thank Dr. Antonia Manova for technical assistance and Kurt Steiner for support in database search. This work was supported by the Swiss National Science Foundation and the Kanton Zurich.

Figure legends

Figure 1: Model transcription factors used in the study.

Chimeric transcription factors, driven by metal-inducible *Drosophila* metallothionein B promoter (pMtnB) or by the alcohol dehydrogenase 1 promoter (pADH1); 20Q, 40Q and 86Q are polyglutamine stretches encoded by CAG codons. 20 and 86 AGC codons encode 20 and 86 serines (20S and 86S).

Figure 2. PolyQs contribute to transcriptional activation in a length-dependent manner in cell culture, in whole flies and in yeast.

A. *Drosophila* Kc cells were transfected with 20 μg of reporter plasmid, 5 μg reference plasmid and 10 μg of effector plasmid. The expression of effector genes (*pMtnB-Gal4DBD-nQ-VP16AD80*) was induced by adding 100 μM $CuSO_4$ to cell growth media for 48 hours. Transcript levels of the 5 x Gal-OVEC reporter and actin-OVEC reference were determined by S1 nuclease protection assay.

B. PolyQ boosts reporter gene expression in transgenic flies.

Fly genotype: $y\ w; \frac{\text{UAS-GFP}}{+}; \frac{\text{pMtnB-Gal4DBD-nQ-VP16AD}^{80}}{+}$

Newly eclosed flies (1-1.5 hour old) were chloroform anesthetized and the pictures of four fly abdomens were taken with the same exposure time (1200 ms). The length of the polyQ stretch (n) in the chimeric transcription factor is indicated; ΔQ = no glutamines.

C. Quantitative beta-galactosidase assay in *Drosophila* larvae.

The fly genotype was as follows: $y\ w; ; \frac{\text{pMtnB-Gal4DBD-nQ-VP16AD}^{80}}{\text{UAS-LacZ}}$

Beta-galactosidase activity was measured in pools of three 3rd instar larvae by the liquid LacZ (ONPG) assay. The value without polyglutamine (ΔQ) was taken as 1.

D. Activity of chimeric transcription factors in yeast *in vivo* plate assay using X-gal in the medium. The increasing blue color intensity of yeast cells indicates increasing activity of the transcription factor that drives the expression of the *LacZ* reporter gene.

E. Quantitative beta-galactosidase assay in yeast liquid culture using ONPG as substrate. The activity of the ΔQ is set to 1.

Figure 3. Polyserine tracts do not contribute to transcription activation.

A. Transcript levels of the 5 x Gal-OVEC reporter and CMV-OVEC reference in human embryonic kidney 293 (HEK293) cells were determined by S1 nuclease protection assay. Cells were transfected with 5 μg of pMtnB-Gal4DBD-nQ/nS-VP16AD^{80} effector constructs, 10 μg reporter plasmid and 2 μg reference plasmid. 500 μM $CuSO_4$ was used to induce the MtnB promoter for 24 hours. For each lane, the size of polyQ or polyS tract is indicated.

B. Test of polyglutamines and polyserines in transgenic *Drosophila* larvae.

Fly genotype: *y w*; $\frac{\text{UAS-GFP}}{+}$; $\frac{\text{pMtnB-Gal4DBD-nQ/nS-VP16AD}^{80}}{+}$

Reporter activity was determined by liquid LacZ assay; the relative β-galactosidase activity for ΔQ is set to 1.

Figure 4. Plasmids used for mammalian two-hybrid studies.

All constructs are expressed from the strong enhancer/promoter of human cytomegalovirus ("pCMV"). The size of the polyQ stretches, encoded by CAG triplets, is indicated.

Figure 5. Length-dependent activity of polyQ tracts in a mammalian two-hybrid system.

HEK293T cells were transfected with 4 μg of each of the two-hybrid effector plasmids, 10 μg 5 x Gal-OVEC reporter and 1 μg of CMV-OVEC reference plasmid. Reporter gene transcripts were quantified by S1 nuclease protection assay. The basal level signal with both bait and prey lacking polyQs is set to 1.

Figure 6. Two-hybrid interaction observed *in vivo* is too weak for detection by EMSA.

No interaction could be detected between Gal4-nQ and VP16-nQ. ^{32}P-labeled oligonucleotides containing 2 x UAS sites (DNA-binding sites for Gal4) were incubated with 5 μg of HEK293T cell nuclear extracts.

Table 1. Examples of eukaryotic transcription factors containing homopolymeric glutamine tracts.

Table 2. Homopolymeric amino acid stretches in the *S. cerevisiae* proteome.

This was done using the Yeast Genome Pattern Matching tool of the Saccharomyces Genome Database [http://www.yeastgenome.org/cgi-bin/PATMATCH/nph-patmatch].

Figure 1.

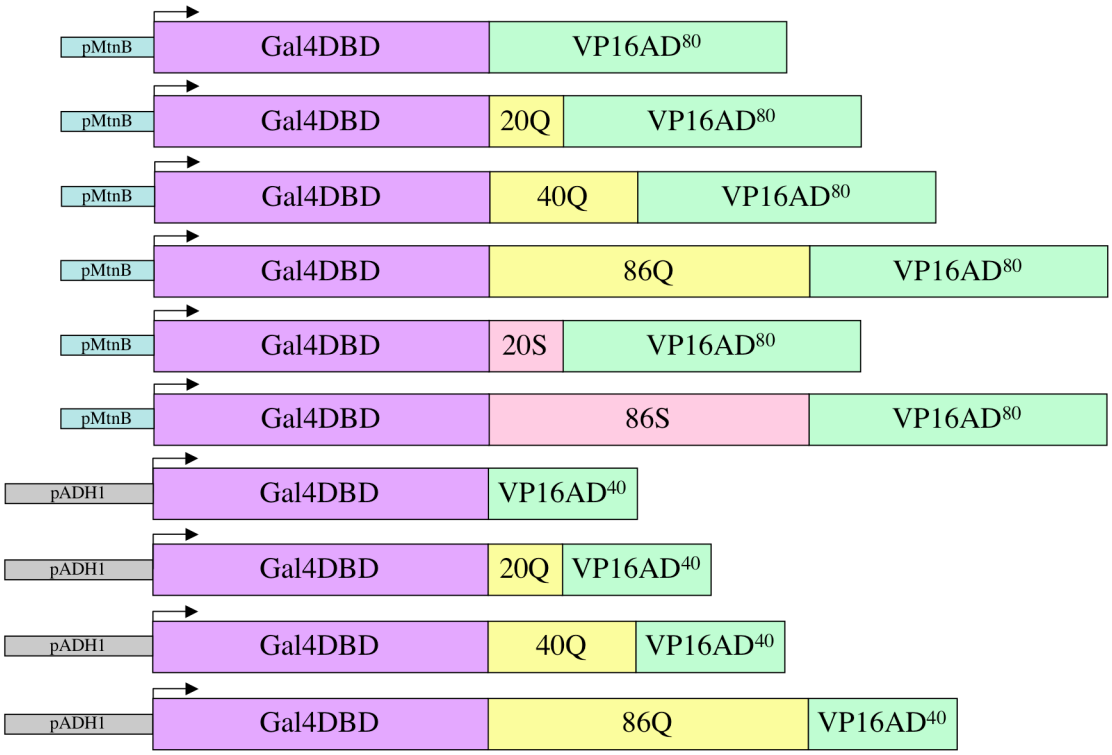
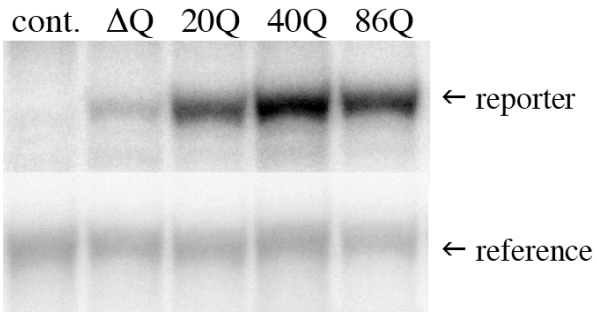
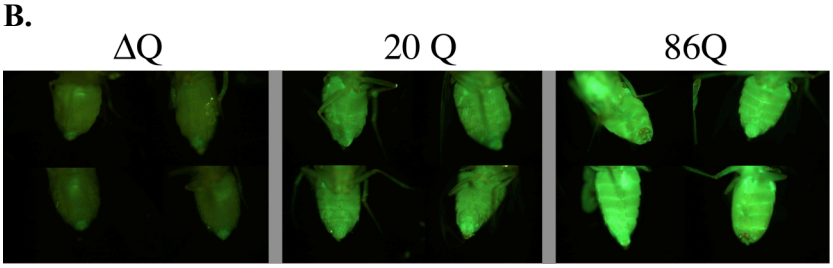


Figure 2.

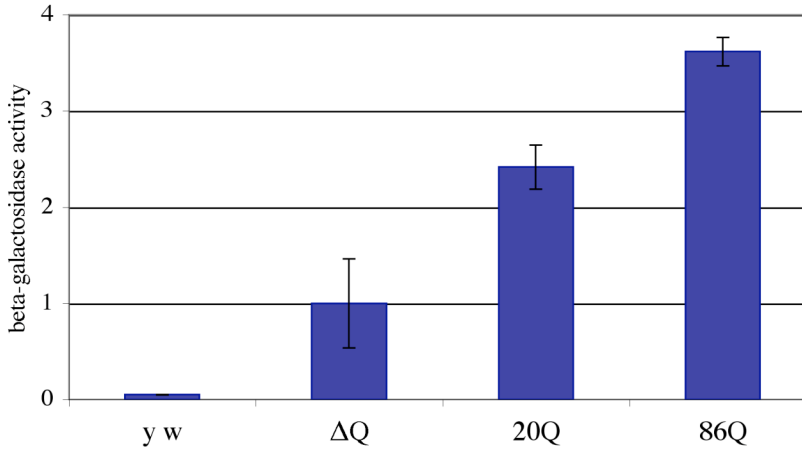
A.



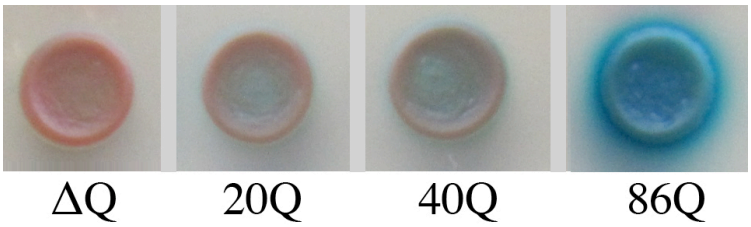
B.



C.


beta-galactosidase activity
4
3
2
1
0
y w
ΔQ
20Q
86Q

D.


ΔQ
20Q
40Q
86Q

E.

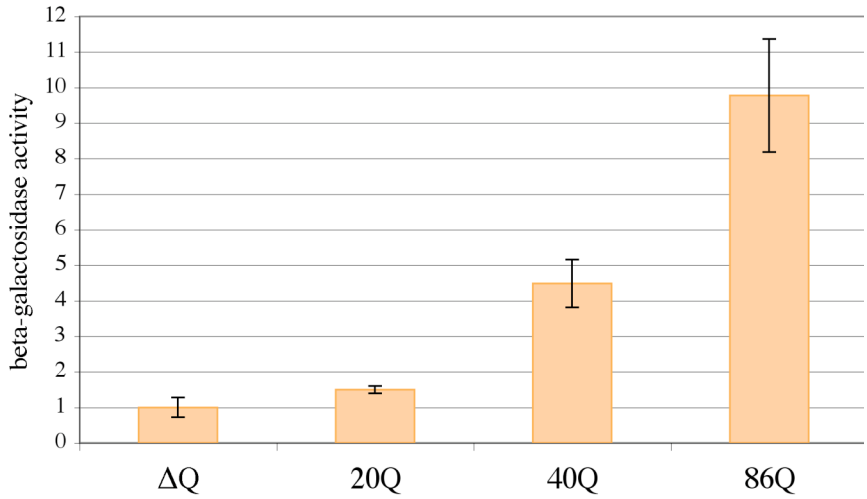
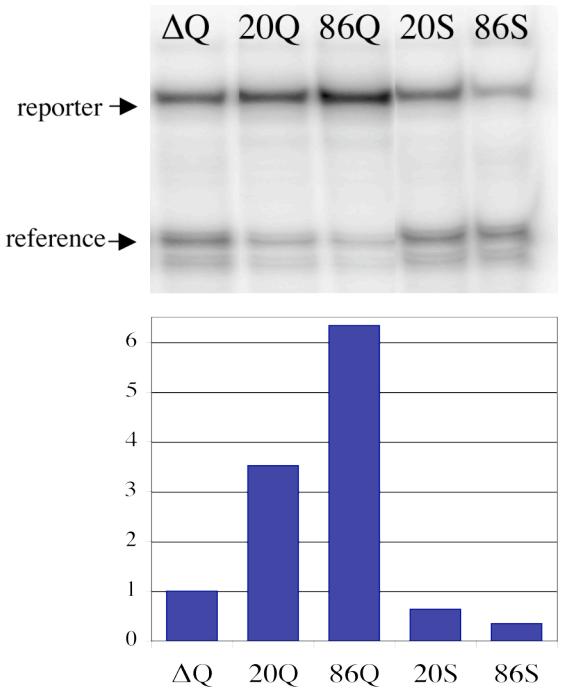

beta-galactosidase activity
12
11
10
9
8
7
6
5
4
3
2
1
0
ΔQ
20Q
40Q
86Q

Figure 3.

A.


ΔQ
20Q
86Q
20S
86S
reporter
reference
6
5
4
3
2
1
0
ΔQ
20Q
86Q
20S
86S

B.

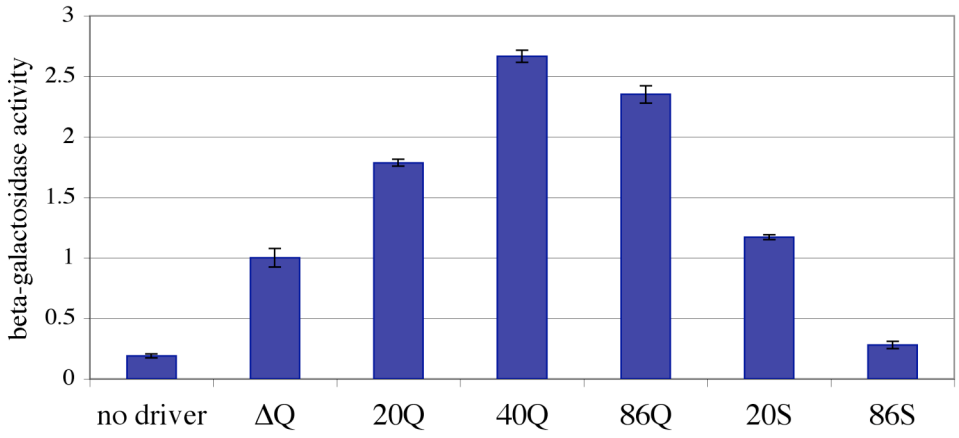

beta-galactosidase activity
3
2.5
2
1.5
1
0.5
0
no driver
ΔQ
20Q
40Q
86Q
20S
86S

Figure 4.

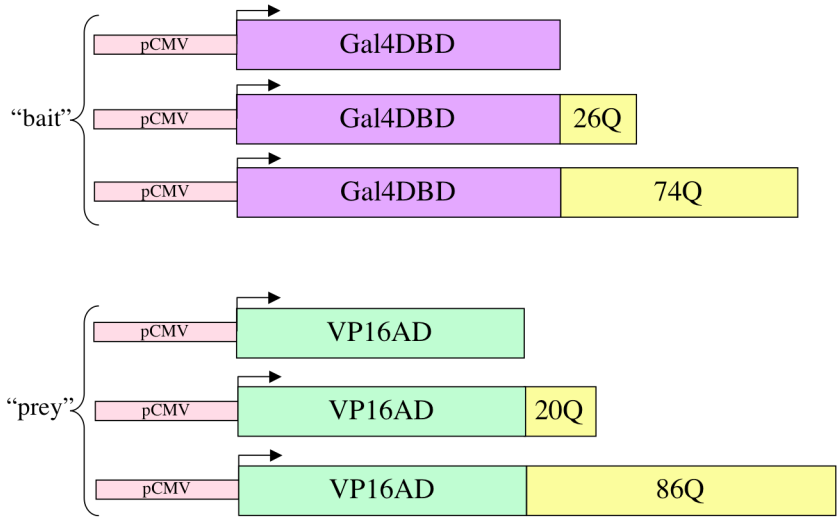


Figure 5.

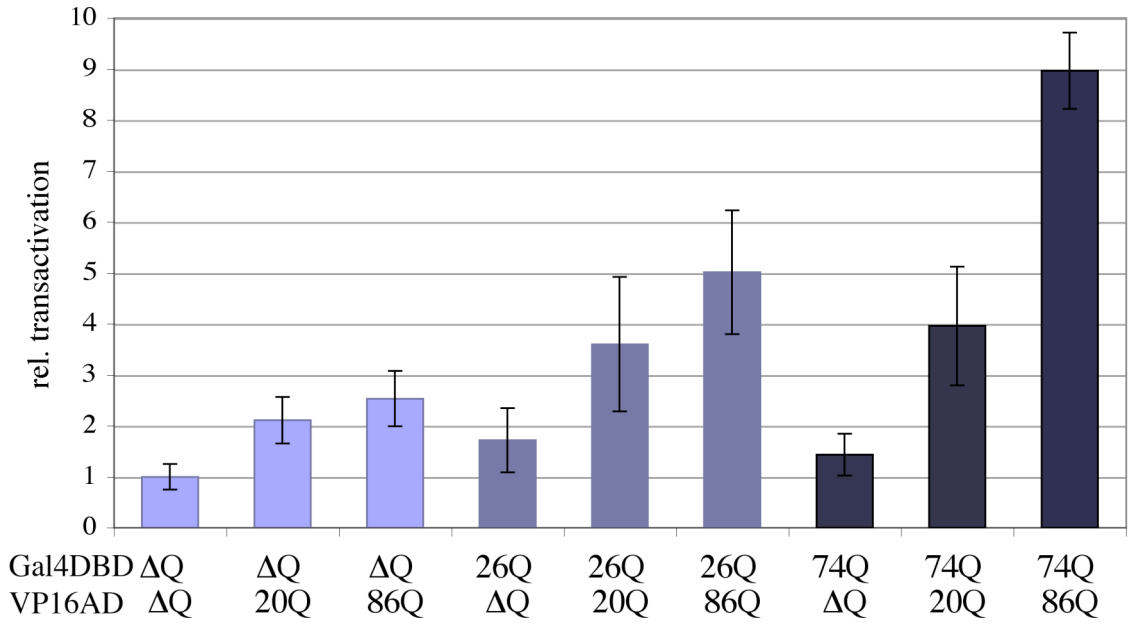


Figure 6.

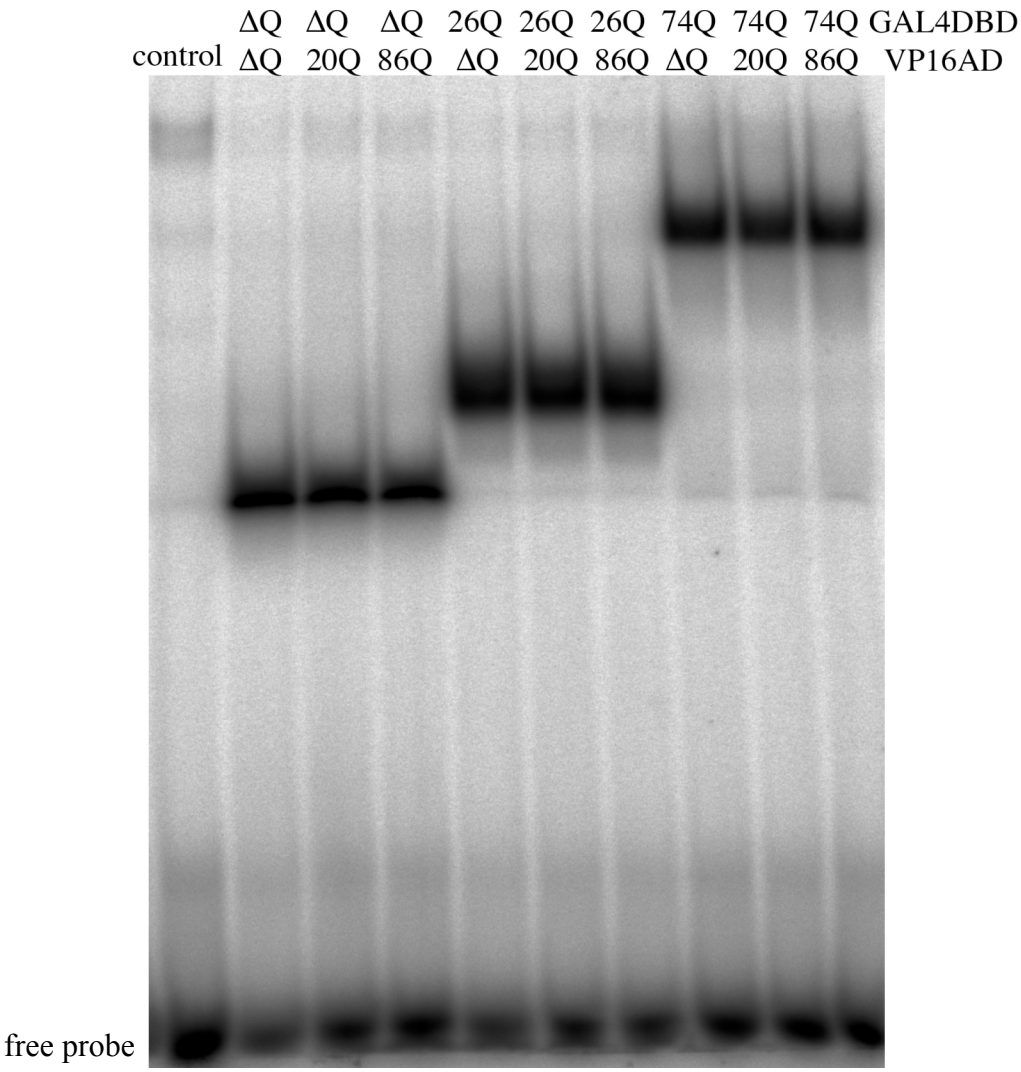

control
ΔQ ΔQ
ΔQ ΔQ
ΔQ 20Q
ΔQ 86Q
26Q ΔQ
26Q 20Q
26Q 86Q
74Q ΔQ
74Q 20Q
74Q 86Q
GAL4DBD VP16AD
free probe

Table 1.

	Transcription factors	Description
Saccaromyces cerevisiae	**Snf5**	Subunit of the SWI/SNF chromatin remodeling complex; is required for expression of a wide variety of differently regulated genes. [Laurent, 1990; Abrams, 1986; Sansam, 2006]
	Mss11	Plays central role in the control of filamentous growth and starch metabolism; regulates transcription of STA2 and MUC1 in response to nutritional signals. [Webber, 1997; Gagiano, 1999; Gagiano, 2003]
	Flo8	Regulates flocculation, diploid filamentous growth, and haploid invasive growth via the transcriptional regulation of the FLO11, STA1 and FLO1 genes. [Kobayashi, 1996; Liu, 1996; Kobayashi, 1999]
	Crz1	"Functions downstream of calceneurin to effect the expression of several stress response genes which confer tolerance to high Ca^{2+}, Mn^{2+}, Na^{+}, cell wall damage. [Stathopoulos, 1997; Matheos, 1997]
	Dal81	Positive regulator of genes in multiple nitrogen degradation pathways. [Bricmont, 1991]
	Gal11	An auxilliary transcription factor required for full expression of nearly all genes in baker's yeast. [Mizuno, 2003; Fukasawa, 2001]
	Hap2	Subunit of the CCAAT-binding complex, a transcriptional activator, global regulator of respiratory gene expression. [Pinkham, 1987; Mc Nabb 2005]
	Ndd1	Essential for nuclear division; is involved in G2/M transcription through its association with forkhead protein 2. [Loy CJ, 1999; Koranda, 2000]
Drosophila melanoguster	**Abdominal-A**	Is required for development of 2-8 abdominal segments. [Karch, 1990; Macias, 1990]
	Suppressor of Hairless (SuH)	Acts as a transciptional activator in the Notch pathway. [Schweisguth, 1992; Bray, 2001]
	Buttonhead	Required for the development of the antennal, intercalary and mandibular segments of the head. [Wimmer, 1993; Wimmer, 1996]
	dCLOCK	Circadian regulator that generates a rhythmic output with a period of about 24 hours. Drives expression of period and timeless which act as inhibitors for dCLOCK. [Allada, 1998; Darlington, 1998]
	CCAAT/enhancer-binding protein	Tissue-specific transcription factor required for embryonic development; it is required for expression of gene products mediating border cell migration [Rørth, 1992; Montell, 1992]
	Prospero	Required for proper neuronal differentiation of most or all neurons and their precursors in central and peripheral nervous systems. [Vaessin, 1991; Chu-Lagraff, 1991]
	GAGA	Regulates expression of homeotic and many other genes. It functions by promoting the open chromatin conformation in promoter regions, allowing access to other transcription factors. [Vaquero, 2000; Katokhin, 2001]
Homo sapiens	**Androgen receptor**	Regulates genes that are critical for the development and maintenance of the male sexual phenotype. [LuBahn, 1988; Sleddens, 1992]
	FOXP2	Is involved in the developmental control of the central nervous sytem, linked to language and speech. [Vargha-Khadem, 2005; Fisher, 2009; Bruce, 2002]
	TATA-box-binding protein (TBP)	General transcription factor that functions at the core of the DNA-binding multiprotein factor TFIID. [van Roon-Mom, 2005; Pugh, 2000; Verrijzer, 1996; Burley, 1996]
	CREB-binding protein (CBP)	Regulates transcription of genes involved in various cellular processes, including intermediary metabolsim, neuronal signalling, cell proliferation and apoptosis. [Mayr, 2001; Mc Manus, 2001]
	Mastermind-like protein 2 and 3	Act as transcriptional coactivators for NOTCH receptors. [Lin, 2002; Wu, 2002]
	Runt-related TF2 (CBFA1)	Plays central role in osteoblastic differentiation and skeletal morphogenesis. [Xiao, 1998; Komori, 1998; Lian, 2003]
	N-Oct-3	Is widely expressed in the developing mammalian central nervious system and is important for neural cell differentiation. [Alazard, 2005; Schreiber, 1993]

Table 2.

amino acid (aa)		> 5 aa	> 8 aa	> 10 aa	> 15 aa	> 20 aa	codons			
Glutamine	**Q**	150	76	51	20	8	CAA	CAG		
Serine	**S**	207	34	11	6	2	AGT	AGC		
Glutamic acid	**E**	105	26	10	3	1	GAA	GAG		
Aspartic acid	**D**	102	33	16	5	4	GAT	GAC		
Asparagine	**N**	101	38	27	11	3	AAT	AAC		
Lysine	**K**	59	4	1	-	-	AAA	AAG		
Proline	**P**	41	8	2	-	-	CCT	CCC	CCA	CCG
Threonine	**T**	35	1	-	-	-	ACT	ACC	ACA	ACG
Alanine	**A**	33	11	-	-	-	GCT	GCC	GCA	GCG
Leucine	**L**	27	-	-	-	-	CTT	CTC	CTA	CTG
Histidine	**H**	16	5	1	-	-	CAT	CAC		
Arginine	**R**	12	1	1	-	-	AGA	AGG		
Glycine	**G**	7	-	-	-	-	GGT	GGC	GGA	GGG
Isoleucine	**I**	5	-	-	-	-	ATT	ATC	ATA	
Phenylalanine	**F**	3	-	-	-	-	TTT	TTC		
Valine	**V**	2	1	-	-	-	GTT	GTC	GTA	GTG
Tyrosine	**Y**	2	-	-	-	-	TAT	TAC		
Tryptophan	**W**	-	-	-	-	-	TGA	TGG		
Methionine	**M**	-	-	-	-	-	ATG			
Cysteine	**C**	-	-	-	-	-	TGT	TGC		

References

1. **Abrams, E., L. Neigeborn, and M. Carlson.** 1986. Molecular analysis of SNF2 and SNF5, genes required for expression of glucose-repressible genes in *Saccharomyces cerevisiae*. Mol Cell Biol **6:**3643-51.

2. **Alazard, R., M. Blaud, S. Elbaz, C. Vossen, G. Icre, G. Joseph, L. Nieto, and M. Erard.** 2005. Identification of the 'NORE' (N-Oct-3 responsive element), a novel structural motif and composite element. Nucleic Acids Res **33:**1513-23.

3. **Allada, R., N. White, W. So, J. Hall, and M. Rosbash.** 1998. A mutant Drosophila homolog of mammalian Clock disrupts circadian rhythms and transcription of period and timeless. Cell **93:**791-804.

4. **Arrasate, M., S. Mitra, E. Schweitzer, M. Segal, and S. Finkbeiner.** 2004. Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. Nature **431:**805-10.

5. **Bettencourt, B., C. Hogan, and M. Nimali.** 2007. Polyglutamine expansion in Drosophila: thermal stress and Hsp70 as selective agents. J Biosci **32:**537-47.

6. **Bischof, J., R. Maeda, M. Hediger, F. Karch, and K. Basler.** 2007. An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. Proc Natl Acad Sci U S A **104:**3312-7.

7. **Bray, S., and M. Furriols.** 2001. Notch pathway: making sense of suppressor of hairless. Curr Biol **11:**R217-21.

8. **Bricmont, P., J. Daugherty, and T. Cooper.** 1991. The DAL81 gene product is required for induced expression of two differently regulated nitrogen catabolic genes in Saccharomyces cerevisiae. Mol Cell Biol **11:**1161-6.

9. **Bruce, H., and R. Margolis.** 2002. FOXP2: novel exons, splice variants, and CAG repeat length stability. Hum Genet **111:**136-44.

10. **Burley, S.** 1996. The TATA box binding protein. Curr Opin Struct Biol **6:**69-75.

11. **Chu-Lagraff, Q., D. Wright, L. McNeil, and C. Doe.** 1991. The prospero gene encodes a divergent homeodomain protein that controls neuronal identity in Drosophila. Development **Suppl 2:**79-85.

12. **Darlington, T., K. Wager-Smith, M. Ceriani, D. Staknis, N. Gekakis, T. Steeves, C. Weitz, J. Takahashi, and S. Kay.** 1998. Closing the circadian loop: CLOCK-induced transcription of its own inhibitors per and tim. Science **280:**1599-603.

13. **Dunah, A., H. Jeong, A. Griffin, Y. Kim, D. Standaert, S. Hersch, M. Mouradian, A. Young, N. Tanese, and D. Krainc.** 2002. Sp1 and TAFII130 transcriptional activity disrupted in early Huntington's disease. Science **296:**2238-43.

14. **Fish, M., A. Groth, M. Calos, and R. Nusse.** 2007. Creating transgenic Drosophila by microinjecting the site-specific phiC31 integrase mRNA and a transgene-containing donor plasmid. Nat Protoc **2:**2325-31.

15. **Fisher, S., and C. Scharff.** 2009. FOXP2 as a molecular window into speech and language. Trends Genet **25:**166-77.

16. **Freiman, R., and R. Tjian.** 2002. Neurodegeneration. A glutamine-rich trail leads to transcription factors. Science **296:**2149-50.

17. **Fujikake, N., Y. Nagai, H. Popiel, Y. Okamoto, M. Yamaguchi, and T. Toda.** 2008. Heat shock transcription factor 1-activating compounds suppress polyglutamine-induced neurodegeneration through induction of multiple molecular chaperones. J Biol Chem **283:**26188-97.

18. **Fukasawa, T., M. Fukuma, K. Yano, and H. Sakurai.** 2001. A genome-wide analysis of transcriptional effect of Gal11 in Saccharomyces cerevisiae: an application of "mini-array hybridization technique". DNA Res **8:**23-31.

19. **Gagiano, M., M. Bester, D. van Dyk, J. Franken, F. Bauer, and I. Pretorius.** 2003. Mss11p is a transcription factor regulating pseudohyphal differentiation, invasive growth and starch metabolism in Saccharomyces cerevisiae in response to nutrient availability. Mol Microbiol **47:**119-34.

20. **Gagiano, M., D. Van Dyk, F. Bauer, M. Lambrechts, and I. Pretorius.** 1999. Divergent regulation of the evolutionarily closely related promoters of the Saccharomyces cerevisiae STA2 and MUC1 genes. J Bacteriol **181:**6497-508.

21. **Gao, R., T. Matsuura, M. Coolbaugh, C. Zühlke, K. Nakamura, A. Rasmussen, M. Siciliano, T. Ashizawa, and X. Lin.** 2008. Instability of expanded CAG/CAA repeats in spinocerebellar ataxia type 17. Eur J Hum Genet **16:**215-22.

22. **Gerber, H., K. Seipel, O. Georgiev, M. Höfferer, M. Hug, S. Rusconi, and W. Schaffner.** 1994. Transcriptional activation modulated by homopolymeric glutamine and proline stretches. Science **263:**808-11.

23. **Graham, F. L., and A. J. van der Eb.** 1973. A new technique for the assay of infectivity of human adenovirus 5 DNA. Virology **52:**456-467.

24. **Gusella, J. F., and M. E. MacDonald.** 2000. Molecular genetics: unmasking polyglutamine triggers in neurodegenerative disease. Nat Rev Neurosci **1:**109-115.

25. **Helmlinger, D., L. Tora, and D. Devys.** 2006. Transcriptional alterations and chromatin remodeling in polyglutamine diseases. Trends Genet **22:**562-70.

26. **Karch, F., W. Bender, and B. Weiffenbach.** 1990. abdA expression in Drosophila embryos. Genes Dev **4:**1573-87.

27. **Katokhin, A., A. Pindiurin, E. Fedorova, and E. Baricheva.** 2001. [Molecular genetic analysis of Thrithorax-like gene encoded transcriptional factor GAGA in Drosophila melanogaster]. Genetika **37:**467-74.

28. **Kazantsev, A., E. Preisinger, A. Dranovsky, D. Goldgaber, and D. Housman.** 1999. Insoluble detergent-resistant aggregates form between pathological and nonpathological lengths of polyglutamine in mammalian cells. Proc Natl Acad Sci U S A **96:**11404-9.

29. **Kobayashi, O., H. Suda, T. Ohtani, and H. Sone.** 1996. Molecular cloning and analysis of the dominant flocculation gene FLO8 from Saccharomyces cerevisiae. Mol Gen Genet **251:**707-15.

30. **Kobayashi, O., H. Yoshimoto, and H. Sone.** 1999. Analysis of the genes activated by the FLO8 gene in Saccharomyces cerevisiae. Curr Genet **36:**256-61.

31. **Komori, T., and T. Kishimoto.** 1998. Cbfa1 in bone development. Curr Opin Genet Dev **8:**494-9.

32. **Koranda, M., A. Schleiffer, L. Endler, and G. Ammerer.** 2000. Forkhead-like transcription factors recruit Ndd1 to the chromatin of G2/M-specific promoters. Nature **406:**94-8.

33. **Kozlowski, P., M. de Mezer, and W. Krzyzosiak.** 2010. Trinucleotide repeats in human genome and exome. Nucleic Acids Res **38:**4027-39.

34. **Krobitsch, S., and S. Lindquist.** 2000. Aggregation of huntingtin in yeast varies with the length of the polyglutamine expansion and the expression of chaperone proteins. Proc Natl Acad Sci U S A **97:**1589-94.

35. **La Spada, A., H. Paulson, and K. Fischbeck.** 1994. Trinucleotide repeat expansion in neurological disease. Ann Neurol **36:**814-22.

36. **Lanz, R., S. Wieland, M. Hug, and S. Rusconi.** 1995. A transcriptional repressor obtained by alternative translation of a trinucleotide repeat. Nucleic Acids Res **23:**138-45.

37. **Laurent, B., M. Treitel, and M. Carlson.** 1990. The SNF5 protein of Saccharomyces cerevisiae is a glutamine- and proline-rich transcriptional activator that affects expression of a broad spectrum of genes. Mol Cell Biol **10:**5616-25.

38. **Legendre, M., N. Pochet, T. Pak, and K. Verstrepen.** 2007. Sequence-based estimation of minisatellite and microsatellite repeat variability. Genome Res **17:**1787-96.

39. **Lian, J., and G. Stein.** 2003. Runx2/Cbfa1: a multifunctional regulator of bone formation. Curr Pharm Des **9:**2677-85.

40. **Lin, S., T. Oyama, T. Nagase, K. Harigaya, and M. Kitagawa.** 2002. Identification of new human mastermind proteins defines a family that consists of positive regulators for notch signaling. J Biol Chem **277:**50612-20.

41. **Liu, H., C. Styles, and G. Fink.** 1996. Saccharomyces cerevisiae S288C has a mutation in FLO8, a gene required for filamentous growth. Genetics **144:**967-78.

42. **Loy, C., D. Lydall, and U. Surana.** 1999. NDD1, a high-dosage suppressor of cdc28-1N, is essential for expression of a subset of late-S-phase-specific genes in Saccharomyces cerevisiae. Mol Cell Biol **19:**3312-27.

43. **Lubahn, D., D. Joseph, M. Sar, J. Tan, H. Higgs, R. Larson, F. French, and E. Wilson.** 1988. The human androgen receptor: complementary deoxyribonucleic acid cloning, sequence analysis and gene expression in prostate. Mol Endocrinol **2:**1265-75.

44. **Macias, A., J. Casanova, and G. Morata.** 1990. Expression and regulation of the abd-A gene of Drosophila. Development **110:**1197-207.

45. **Matheos, D., T. Kingsbury, U. Ahsan, and K. Cunningham.** 1997. Tcn1p/Crz1p, a calcineurin-dependent transcription factor that differentially regulates gene expression in Saccharomyces cerevisiae. Genes Dev **11:**3445-58.

46. **Mayr, B., and M. Montminy.** 2001. Transcriptional regulation by the phosphorylation-dependent factor CREB. Nat Rev Mol Cell Biol **2:**599-609.

47. **McManus, K., and M. Hendzel.** 2001. CBP, a transcriptional coactivator and acetyltransferase. Biochem Cell Biol **79:**253-66.

48. **McNabb, D., and I. Pinto.** 2005. Assembly of the Hap2p/Hap3p/Hap4p/Hap5p-DNA complex in Saccharomyces cerevisiae. Eukaryot Cell **4:**1829-39.

49. **Merlino, A., L. Esposito, and L. Vitagliano.** 2006. Polyglutamine repeats and beta-helix structure: molecular dynamics study. Proteins **63:**918-927.

50. **Miller, J., M. Arrasate, B. Shaby, S. Mitra, E. Masliah, and S. Finkbeiner.** 2010. Quantitative relationships between huntingtin levels, polyglutamine length, inclusion body formation, and neuronal death provide novel insight into huntington's disease molecular pathogenesis. J Neurosci **30:**10541-50.

51. **Mirkin, S.** 2007. Expandable DNA repeats and human disease. Nature **447:**932-40.

52. **Mizuno, T., and S. Harashima.** 2003. Gal11 is a general activator of basal transcription, whose activity is regulated by the general repressor Sin4 in yeast. Mol Genet Genomics **269:**68-77.

53. **Montell, D., P. Rorth, and A. Spradling.** 1992. slow border cells, a locus required for a developmentally regulated cell migration during oogenesis, encodes Drosophila C/EBP. Cell **71:**51-62.

54. **Neef, D., M. Turski, and D. Thiele.** 2010. Modulation of heat shock transcription factor 1 as a therapeutic target for small molecule intervention in neurodegenerative disease. PLoS Biol **8:**e1000291.

55. **Nucifora , F. J., M. Sasaki, M. Peters, H. Huang, J. Cooper, M. Yamada, H. Takahashi, S. Tsuji, J. Troncoso, V. Dawson, T. Dawson, and C. Ross.** 2001. Interference by huntingtin and atrophin-1 with cbp-mediated transcription leading to cellular toxicity. Science **291:**2423-8.

56. **Okazawa, H.** 2003. Polyglutamine diseases: a transcription disorder? Cell Mol Life Sci **60:**1427-39.

57. **Oma, Y., Y. Kino, K. Toriumi, N. Sasagawa, and S. Ishiura.** 2007. Interactions between homopolymeric amino acids (HPAAs). Protein Sci **16:**2195-204.

58. **Orr, H., and H. Zoghbi.** 2007. Trinucleotide repeat disorders. Annu Rev Neurosci **30:**575-621.

59. **Perutz, M., J. Finch, J. Berriman, and A. Lesk.** 2002. Amyloid fibers are water-filled nanotubes. Proc Natl Acad Sci U S A **99:**5591-5.

60. **Perutz, M., T. Johnson, M. Suzuki, and J. Finch.** 1994. Glutamine repeats as polar zippers: their possible role in inherited neurodegenerative diseases. Proc Natl Acad Sci U S A **91:**5355-8.

61. **Pinkham, J., J. Olesen, and L. Guarente.** 1987. Sequence and nuclear localization of the Saccharomyces cerevisiae HAP2 protein, a transcriptional activator. Mol Cell Biol **7:**578-85.

62. **Poirier, M., H. Li, J. Macosko, S. Cai, M. Amzel, and C. Ross.** 2002. Huntingtin spheroids and protofibrils as precursors in polyglutamine fibrilization. J Biol Chem **277:**41032-7.

63. **Pugh, B.** 2000. Control of gene expression through regulation of the TATA-binding protein. Gene **255:**1-14.

64. **Riley, B., and H. Orr.** 2006. Polyglutamine neurodegenerative diseases and regulation of transcription: assembling the puzzle. Genes Dev **20:**2183-92.

65. **Robertson, A., S. Headey, H. Saunders, H. Ecroyd, M. Scanlon, J. Carver, and S. Bottomley.** 2010. Small heat-shock proteins interact with a flanking domain to suppress polyglutamine aggregation. Proc Natl Acad Sci U S A **107:**10424-9.

66. **Rørth, P., and D. Montell.** 1992. Drosophila C/EBP: a tissue-specific DNA-binding protein required for embryonic development. Genes Dev **6:**2299-311.

67. **Sansam, C., and C. Roberts.** 2006. Epigenetics and cancer: altered chromatin remodeling via Snf5 loss leads to aberrant cell cycle regulation. Cell Cycle **5:**621-4.

68. **Schaffar, G., P. Breuer, R. Boteva, C. Behrends, N. Tzvetkov, N. Strippel, H. Sakahira, K. Siegers, M. Hayer-Hartl, and F. Hartl.** 2004. Cellular toxicity of polyglutamine expansion proteins: mechanism of transcription factor deactivation. Mol Cell **15:**95-105.

69. **Scherzinger, E., A. Sittler, K. Schweiger, V. Heiser, R. Lurz, R. Hasenbank, G. Bates, H. Lehrach, and E. Wanker.** 1999. Self-assembly of polyglutamine-containing huntingtin fragments into amyloid-like fibrils: implications for Huntington's disease pathology. Proc Natl Acad Sci U S A **96:**4604-9.

70. **Schreiber, E., A. Tobler, U. Malipiero, W. Schaffner, and A. Fontana.** 1993. cDNA cloning of human N-Oct3, a nervous-system specific POU domain transcription factor binding to the octamer DNA motif. Nucleic Acids Res **21:**253-8.

71. **Schweisguth, F., and J. Posakony.** 1992. Suppressor of Hairless, the Drosophila homolog of the mouse recombination signal-binding protein gene, controls sensory organ cell fates. Cell **69:**1199-212.

72. **Seipel, K., O. Georgiev, and W. Schaffner.** 1992. Different activation domains stimulate transcription from remote ('enhancer') and proximal ('promoter') positions. EMBO J **11:**4961-8.

73. **Shimohata, T., T. Nakajima, M. Yamada, C. Uchida, O. Onodera, S. Naruse, T. Kimura, R. Koide, K. Nozaki, Y. Sano, H. Ishiguro, K. Sakoe, T. Ooshima, A. Sato, T. Ikeuchi, M. Oyake, T. Sato, Y. Aoyagi, I. Hozumi, T. Nagatsu, Y. Takiyama, M. Nishizawa, J. Goto, I. Kanazawa, I. Davidson, N. Tanese, H. Takahashi, and S. Tsuji.** 2000. Expanded polyglutamine stretches interact with TAFII130, interfering with CREB-dependent transcription. Nat Genet **26:**29-36.

74. **Sleddens, H., B. Oostra, A. Brinkmann, and J. Trapman.** 1992. Trinucleotide repeat polymorphism in the androgen receptor gene (AR). Nucleic Acids Res **20:**1427.

75. **Slow, E., R. Graham, A. Osmand, R. Devon, G. Lu, Y. Deng, J. Pearson, K. Vaid, N. Bissada, R. Wetzel, B. Leavitt, and M. Hayden.** 2005. Absence of behavioral abnormalities and neurodegeneration in vivo despite widespread neuronal huntingtin inclusions. Proc Natl Acad Sci U S A **102:**11402-7.

76. **Stathopoulos, A., and M. Cyert.** 1997. Calcineurin acts through the CRZ1/TCN1-encoded transcription factor to regulate gene expression in yeast. Genes Dev **11:**3432-44.

77. **Sugars, K., and D. Rubinsztein.** 2003. Transcriptional abnormalities in Huntington disease. Trends Genet **19:**233-8.

78. **Sánchez, I., C. Mahlke, and J. Yuan.** 2003. Pivotal role of oligomerization in expanded polyglutamine neurodegenerative disorders. Nature **421:**373-9.

79. **Tanese, N., and R. Tjian.** 1993. Coactivators and TAFs: a new class of eukaryotic transcription factors that connect activators to the basal machinery. Cold Spring Harb Symp Quant Biol **58:**179-85.

80. **Tasset, D., L. Tora, C. Fromental, E. Scheer, and P. Chambon.** 1990. Distinct classes of transcriptional activating domains function by different mechanisms. Cell **62:**1177-87.

81. **Vaessin, H., E. Grell, E. Wolff, E. Bier, L. Jan, and Y. Jan.** 1991. prospero is expressed in neuronal precursors and encodes a nuclear protein that is involved in the control of axonal outgrowth in Drosophila. Cell **67:**941-53.

82. **van Roon-Mom, W., S. Reid, R. Faull, and R. Snell.** 2005. TATA-binding protein in neurodegenerative disease. Neuroscience **133:**863-72.

83. **Vaquero, A., M. Espinás, F. Azorin, and J. Bernueś.** 2000. Functional mapping of the GAGA factor assigns its transcriptional activity to the C-terminal glutamine-rich domain. J Biol Chem **275:**19461-8.

84. **Vargha-Khadem, F., D. Gadian, A. Copp, and M. Mishkin.** 2005. FOXP2 and the neuroanatomy of speech and language. Nat Rev Neurosci **6:**131-8.

85. **Verrijzer, C., and R. Tjian.** 1996. TAFs mediate transcriptional activation and promoter selectivity. Trends Biochem Sci **21:**338-42.

86. **Weaver, R., and C. Weissmann.** 1979. Mapping of RNA by a modification of the Berk-Sharp procedure: the 5' termini of 15 S beta-globin mRNA precursor and mature 10 s beta-globin mRNA have identical map coordinates. Nucleic Acids Res **7:**1175-93.

87. **Webb, D. M., and J. Zhang.** 2005. FoxP2 in song-learning birds and vocal-learning mammals. J Hered **96:**212-216.

88. **Webber, A., M. Lambrechts, and I. Pretorius.** 1997. MSS11, a novel yeast gene involved in the regulation of starch metabolism. Curr Genet **32:**260-6.

89. **Westin, G., T. Gerster, M. Müller, G. Schaffner, and W. Schaffner.** 1987. OVEC, a versatile system to study transcription in mammalian cells and cell-free extracts. Nucleic Acids Res **15:**6787-98.

90. **Wimmer, E., G. Frommer, B. Purnell, and H. Jäckle.** 1996. buttonhead and D-Sp1: a novel Drosophila gene pair. Mech Dev **59:**53-62.

91. **Wimmer, E., H. Jäckle, C. Pfeifle, and S. Cohen.** 1993. A Drosophila homologue of human Sp1 is a head-specific segmentation gene. Nature **366:**690-4.

92. **Wu, L., T. Sun, K. Kobayashi, P. Gao, and J. Griffin.** 2002. Identification of a family of mastermind-like transcriptional coactivators for mammalian notch receptors. Mol Cell Biol **22:**7688-700.

93. **Xiao, Z., R. Thomas, T. Hinson, and L. Quarles.** 1998. Genomic structure and isoform expression of the mouse, rat and human Cbfa1/Osf2 transcription factor. Gene **214:**187-97.

94. **Zhang, B., D. Egli, O. Georgiev, and W. Schaffner.** 2001. The Drosophila homolog of mammalian zinc finger factor MTF-1 activates transcription in response to heavy metals. Mol Cell Biol **21:**4505-4514.

Stability/instability of triplet repeats

Relative to non-repeated DNA, tandemly repeated DNA sequences are genetically less stable; in case of triplet repeats, variations in repeat numbers, especially when of repeats present in the protein coding regions, can cause pathological conditions. Such disease-associated proteins include huntingtin, androgen receptor and TATA-binding protein. Instability is intrinsically connected to repeat length, as only tracts above a stability threshold (35-40) tend to become unstable. Repeat instability, often in favour of expansion, also underlines the phenomenon of anticipation, whereby the disease occurs earlier with greater severity in successive generations.

Repeat homogeneity is associated with greater instability, and thus a risk factor for polyQ diseases. For example, in spinocerebellar ataxia 2 (SCA2), CAG repeat tracts with interspersed CAA triplets are common in normal alleles, while expanded alleles usually contain a pure CAG repeat tract [Choudhry, 2001]. In the case of another pathological condition, spinocerebellar ataxia 17 (SCA17), mutation frequency in patients harboring pure CAG repeats is 2-3 fold higher compared to those with CAA interruptions. Interestingly, the pure CAG repeats showed both expansions and deletions while the interrupted repeats exhibited mostly deletions, and these at a significantly lower frequency [Gao, 2008]. Unlike in polyQ-associated neurodegenerative diseases, in the *AIB1* (amplified in breast cancer gene 1) gene polyQ length is stable, probably due to the frequent interruption of poly(CAG) by CAA triplets [Dai, 2003]. The polyQ-encoding region of *AIB1* is unstable in breast cancer cell lines and primary tumours due to point mutations, small deletions or insertions of one or two trinucleotide repeats.

Studies in a *Drosophila* model of human spinocerebellar ataxia 3 (SCA3) over generations showed that CAG repeats were stable with only minimal repeat changes, whereas SCA3 transgenes in germ cells showed high rates of instability [Jung, 2007]. Curiously, in the case of spinocerebellar ataxia type 7 (SCA7), 90 CAG repeats were extremely stable in *Drosophila*, regardless of the context [Jackson, 2005].

Apart from the negative, pathogenic consequences of repeat variability, hypermutable repeats may also have a beneficial role. Variable repeats located in certain key genes,

such as those for transcriptional regulators and chromatin modifiers [Legendre, 2007], makes these genes hypervariable, allowing for fast adaptive evolution [Rando, 2007].

We want to address the question whether CAG repeat size changes can occur spontaneously, or whether a certain selective pressure is required to provoke changes in the CAG repeat size and thus modulate the protein to achieve a desired property. We tested the size of a polyQ tract encoded by pure CAG tracts in a synthetic transcription factor in transgenic *D. melanogaster* strains. This protein is composed of the 1-93 aa of DNA-binding domain of yeast Gal4 transcription factor (Gal4DBD) and the carboxyl-terminal 80 aa of the herpes simplex virus protein VP16 ($VP16AD^{80}$), which are linked via a polyQ stretch: $Gal4DBD\text{-}polyQ\text{-}VP16AD^{80}$. The initial size of polyQ tract was 20, 40 and 86 glutamines, all encoded by CAG triplets.

The size of the poly(CAG) region encoding for polyQ tract was determined by PCR 24 and 60 generations after we obtained the transgenic *D. melanogaster*. Interestingly, we did not observe any gross changes of the size of polyQ-encoding region, even in the case of the longest — 86 CAG repeats (Figure 1).

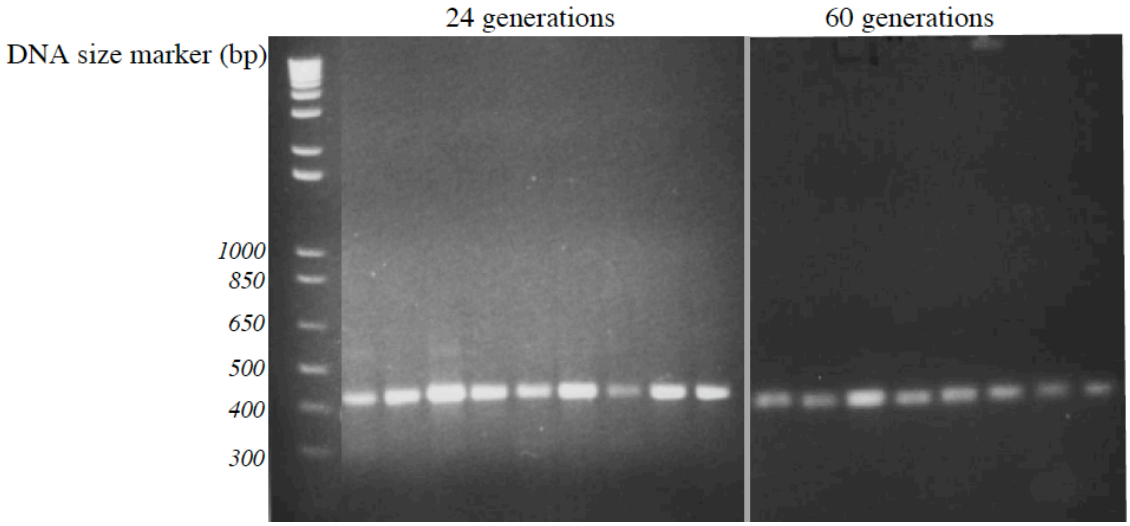


Figure 1. Single fly PCR to amplify a genomic region containing 86 CAG codons.

20-40 transgenic flies (heterozygous for $Gal4DBD\text{-}86Q\text{-}VP16AD^{80}$ transgene) were analyzed. The expected PCR product size for the wild-type transgene is 420 base pairs (bp); the PCR products were analyzed on 2% agarose gel.

The longest physiological polyQ stretch is in FoxP2, a transcription factor involved in the development of the central nervous system, linked to language and speech [Vargha-Khadem, 2005; Fisher, 2009; Bruce, 2002; Konopka, 2009]. *FOXP2* is the first gene that was associated with a language disorder, and it has been subject to positive selection in the human lineage. Mutations in the gene encoding the transcription factor FoxP2 impair human speech and language. FoxP2 is an extremely conserved protein and the long polyQ stretch (~40) is also conserved from Xenopus to humans (Figure 2).

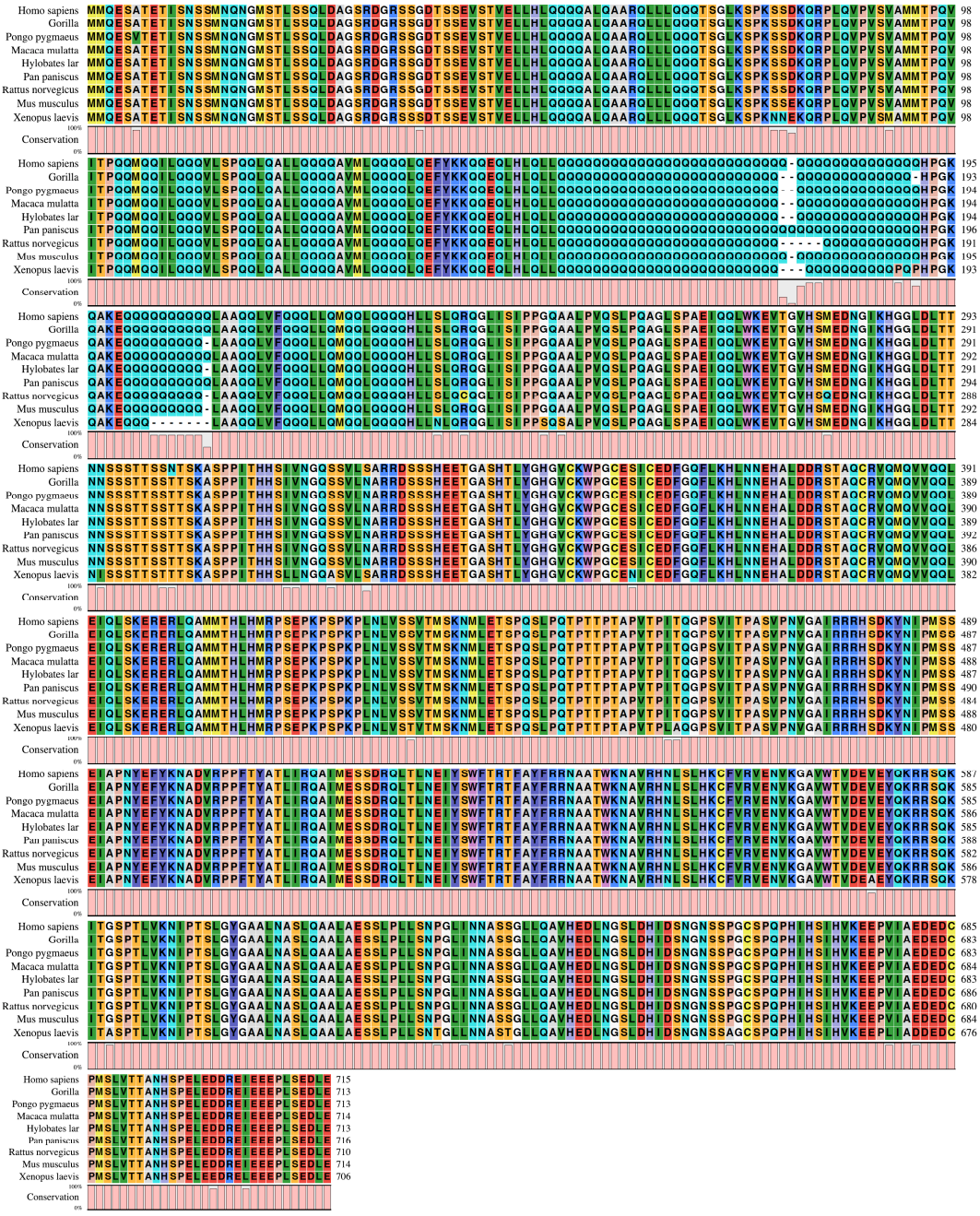


Figure 2. FoxP2 amino acid sequence from different organisms.

All the sequences are derived from SwissProt protein database. The sequence alignments are generated with the CLC Sequence Viewer, Version 6.3, with standard parameters.

We want to test the stability/instability of the polyQ-coding DNA of human FoxP2 (encoded by mixed codons: CAA and CAG) versus pure CAG repeats coding for 40 glutamines (Figure 3). In standard yeast growth conditions, we did not observe polyQ-encoding region instability after 8 generations (data not shown).

a	CAG CAG CAG CAG CAA CAG CAG CAG CAG CAA CAA CAG CAG CAA CAA CAG CAG CAG CAA CAA CAA CAA CAA CAG CAG CAA CAA CAG CAG CAG CAG CAG CAA CAG CAG CAG CAG CAG CAA CAG
b	CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG

Figure 3. Nucleotide sequences encoding for 40 glutamines.

PolyQ-encoding sequence from human FoxP2 gene (**a**) and synthetic pure CAG sequence (**b**).

In the outlook we would like to induce genetic instability into yeast by damaging DNA by UV light, or using such genotoxic agents like cadmium or N-Methyl-N'-Nitro-N-nitrosoguanidine (MNNG) and look at the size and configuration of polyQ-encoding region.

References

1. **Bruce, H., and R. Margolis.** 2002. FOXP2: novel exons, splice variants, and CAG repeat length stability. Hum Genet **111:**136-144.

2. **Choudhry, S., M. Mukerji, A. K. Srivastava, S. Jain, and S. K. Brahmachari.** 2001. CAG repeat instability at SCA2 locus: anchoring CAA interruptions and linked single nucleotide polymorphisms. Hum Mol Genet **10:**2437-2446.

3. **Dai, P., and L. J. Wong.** 2003. Somatic instability of the DNA sequences encoding the polymorphic polyglutamine tract of the AIB1 gene. J Med Genet **40:**885-890.

4. **Fisher, S., and C. Scharff.** 2009. FOXP2 as a molecular window into speech and language. Trends Genet **25:**166-177.

5. **Gao, R., T. Matsuura, M. Coolbaugh, C. Zühlke, K. Nakamura, A. Rasmussen, M. Siciliano, T. Ashizawa, and X. Lin.** 2008. Instability of expanded CAG/CAA repeats in spinocerebellar ataxia type 17. Eur J Hum Genet **16:**215-222.

6. **Jackson, S. M., A. J. Whitworth, J. C. Greene, R. T. Libby, S. L. Baccam, L. J. Pallanck, and A. R. La Spada.** 2005. A SCA7 CAG/CTG repeat expansion is stable in Drosophila melanogaster despite modulation of genomic context and gene dosage. Gene **347:**35-41.

7. **Jung, J., and N. Bonini.** 2007. CREB-binding protein modulates repeat instability in a Drosophila model for polyQ disease. Science **315:**1857-1859.

8. **Konopka, G., J. Bomar, K. Winden, G. Coppola, Z. Jonsson, F. Gao, S. Peng, T. Preuss, J. Wohlschlegel, and D. Geschwind.** 2009. Human-specific transcriptional regulation of CNS development genes by FOXP2. Nature **462:**213-217.

9. **Legendre, M., N. Pochet, T. Pak, and K. Verstrepen.** 2007. Sequence-based estimation of minisatellite and microsatellite repeat variability. Genome Res **17:**1787-1796.

10. **Rando, O. J., and K. J. Verstrepen.** 2007. Timescales of genetic and epigenetic inheritance. Cell **128:**655-668.

11. **Vargha-Khadem, F., D. Gadian, A. Copp, and M. Mishkin.** 2005. FOXP2 and the neuroanatomy of speech and language. Nat Rev Neurosci **6:**131-138.

PART II

Introduction

Regulation of metal homeostasis

Metals are essential for cellular life; they serve as important functional and structural components of many proteins, such as hemoglobin (heme-containing oxygen-transport metalloprotein in the red blood cells), cytochrome C (small heme protein, essential component of the electron transport chain), ceruloplasmin (major copper-carrying protein in the blood), nitrogenase (used by nitrogen-fixing bacteria), superoxide dismutases (a class of enzymes that catalyze the conversion of superoxide into oxygen and hydrogen peroxide). Besides the functionally/structurally essential metals such as iron (Fe), zinc (Zn) and copper (Cu), living organisms can encounter non-essential, highly toxic metals such as mercury (Hg), cadmium (Cd), lead (Pb) and silver (Ag).

In high doses all metal ions are toxic. Problems can be caused basically in two ways: a) by incorrect incorporation into proteins, i.e. displacing correct metal with the abundant metal (mis-metallation), b) by generation of free radicals via the Fenton redox reaction:

$$M^{n+} + H_2O_2 \rightarrow M^{(n+1)+} + OH + OH^-$$

(M stands for a transition metal, especially Fe^{2+} or Cu^{+}).

Some severe congenital diseases are caused by a distorted metal homeostasis. They include Menkes disease, Wilson disease, acrodermatitis enteropathica (see below). Also, several neurodegenerative diseases such as Parkinson's disease, Alzheimer's disease and Huntington's disease are associated with metal imbalances [Rivera-Mancía, 2010].

Living organisms have developed various mechanisms to regulate the intra- and extracellular metal balance. Metal scavengers, metal transporters, metallochaperones and metal regulatory factors function together to achieve metal homeostasis and detoxification (Figure 1).

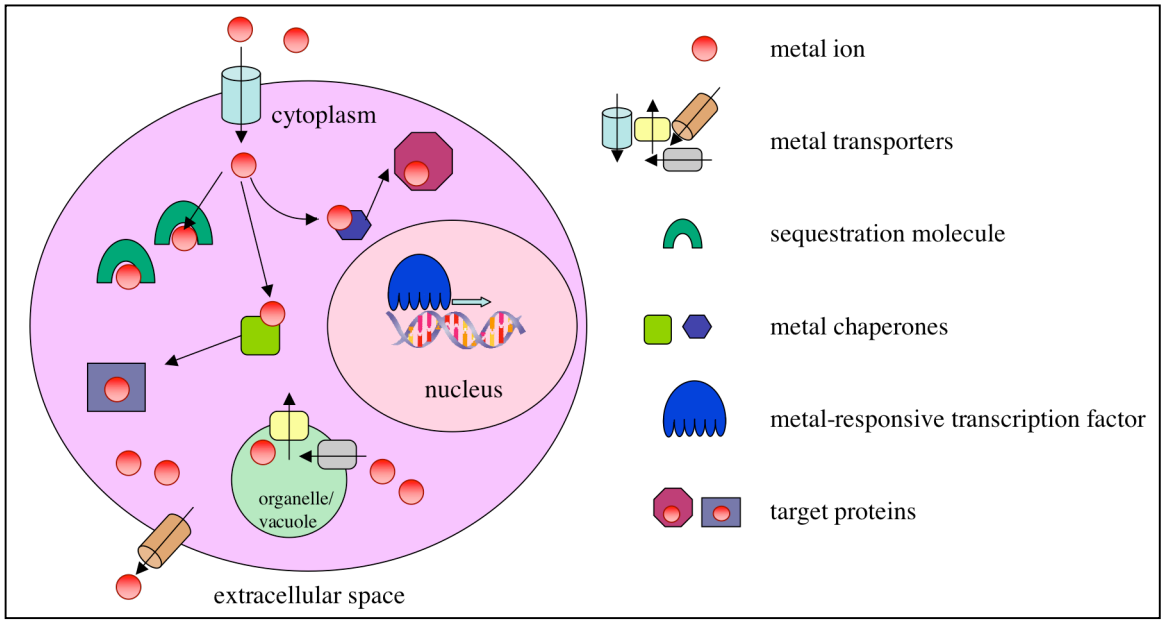


Figure 1. Cellular processes for metal homeostasis.

Various metal importers and exporters are involved in trafficking of metals. Within the cell a metal ion can be handled by sequestration molecules (mostly metallothioneins) or by metal chaperones, which guide the metal ion to the target protein. An important aspect of metal homeostasis is the transcriptional regulation; in response to changes in metal concentrations, metal-responsive transcription factors coordinate the expression of genes that are involved in acquisition, distribution, sequestration and export of metals.

Metal ion transporters play an important role in the metal homeostasis of cells and organisms. Examples of metal transporters are, for copper, ATP7A/MNK, ATP7B/WND and Ctr family transporters; for zinc, Znt and Zip family transporters and for iron, DMT1/Nramp2, SFT transporters. Alterations of the function of metal transporter molecules can result in pathological conditions, for example acrodermatitis enteropathica (AE), an autosomal recessive disease of zinc deficiency results in the mutations in the human zinc transporter gene *hZIP4*, a member of Zip family of zinc transporters [Dufner-Beattie, 2003]. Mutations in ATP7A or ATP7B disrupt the homeostatic copper balance, resulting in Menkes disease or Wilson disease, respectively. Menkes disease or kinky hair disease is characterized by copper accumulation in the kidney and intestine and scarcity in the brain and other organs [Menkes, 1962, de Bie, 2007, Tümer, 2010]. Wilson disease is associated with copper hyperaccumulation in liver and brain [Wilson, 1912; Das, 2006].

Metallochaperones comprise a class of proteins which bind metal ions and deliver them directly to target enzymes via protein-protein interactions [Rosenzweig, 2002]. Examples are mammalian Atox1/Hah1 (with its orthologs Atx1 in yeast, Cch in plants, CopZ in bacteria), which transfers copper to the copper export P-type ATPases ATP7A and ATP7B that are localized in the trans-Golgi network to facilitate copper excretion in eukaryotes [Walker, 2002; Hamza, 2003]; or Ccs1, the copper chaperone for Sod1 conserved from yeast to mammals. Metallochaperones are also believed to deliver nickel ions to bacterial enzymes such as urease and hydrogenase [Stola, 2006; Schauer, 2010]. To date, no specific zinc chaperone has been identified in eukaryotes. A large set of zinc chaperones similar to the dedicated copper chaperones [Robinson, 2010] to serve individual enzymes seems unlikely to exist, considering on the one hand the great diversity of zinc-containing enzymes and transcription factors and on the other hand the considerably lesser toxicity of zinc compared to, for example, copper.

Metal-responsive transcription factors play a key role in regulating the expression of the metal scavengers, metal transporters and metallochaperones. Among these transcription factors are:

I. in yeast:

a) Zap1, zinc-regulated transcription factor; responds to high and low zinc, regulates its own transcription [Zhao, 1997; Lyons, 2000];
b) Aft1, Aft2, which activate genes involved in intracellular iron use and are required for iron homeostasis and resistance to oxidative stress [Yamaguchi-Iwai, 1995; Blaiseau, 2001; Yamaguchi-Iwai, 1996; Shakoury-Elizeh, 2004; Courel, 2005];
c) Fep1, an iron sensor which regulates iron transporter gene expression [Pelletier, 2002; Pelletier, 2005];
d) Ace1/CUP2, a copper-binding transcription factor; it activates transcription of the metallothionein genes *CUP1* and *CRS5* in response to elevated copper concentrations [Thiele, 1988; Buchman, 1989; Culotta, 1994];
e) Mac1, which induces transcription of the *FRE1* and *CTR1* genes upon copper starvation [Jungmann, 1993; Yamaguchi-Iwai, 1997];
f) Cuf1, a copper metalloregulatory transcription factor in fission yeast [Beaudoin, 2001; Beaudoin, 2003].

II. in mammals:

MTF-1, a zinc finger containing transcription factor conserved from mammals to insects. It regulates the expression of metallothioneins, zinc transporters, and several cellular stress response genes [Radtke, 1993; Auf der Maur, 1999; Andrews, 2001a; Lichtlen, 2001; Giedroc, 2001; Zhang, 2001; Yepiskopsyan, 2006].

Metal-responsive transcription factor-1: target genes and interacting proteins

As mentioned, MTF-1 is a transcription factor that regulates gene expression in response to heavy metals. It also has supporting roles in handling oxidative stress and hypoxia [Radtke, 1993; Auf der Maur, 1999; Andrews, 2001a; Lichtlen, 2001; Giedroc, 2001; Zhang, 2001; Yepiskopsyan, 2006]. MTF-1 was discovered in 1988 by Westin and Schaffner and identified as a protein which requires elevated zinc concentrations for optimal DNA binding [Westin, 1988]. Later the corresponding gene was cloned and characterized as a ubiquitously expressed zinc finger transcription factor essential for basal and heavy metal-induced expression of metallothioneins [Radtke, 1993; Brugnera, 1994; Heuchel, 1994].

Mammalian MTF-1 has a molecular weight of approximately 75 kDa (mouse and human MTF-1 are 675 and 753 amino acids long, respectively). The N-terminal domain is followed by six Cys2-His2 type zinc fingers. Mammalian MTF-1 contains three activation domains: an acidic, a proline-rich and a serine/threonine-rich one [Radtke, 1995], as well as nuclear localization and nuclear export signals. A conspicuous cysteine cluster (CQCQCAC) near the C-terminus is required for transcriptional activity [Chen, 2004] and mediates homodimerization of human MTF-1 [V. Günther and W. Schaffner, unpublished data]. Recently it was shown that the nuclear localization signal (NLS) of human MTF-1 is a non-conventional one, spanning zinc fingers 1-3 within the DNA-binding domain [Lindert, 2009].

Upon heavy metal stress, MTF-1 is activated and binds to its DNA sequence motifs called metal response elements (MREs) which are located upstream of its target genes, mainly metallothioneins (MTs), thereby inducing transcription of the cell's major metal-scavenging molecules (Figure 2).

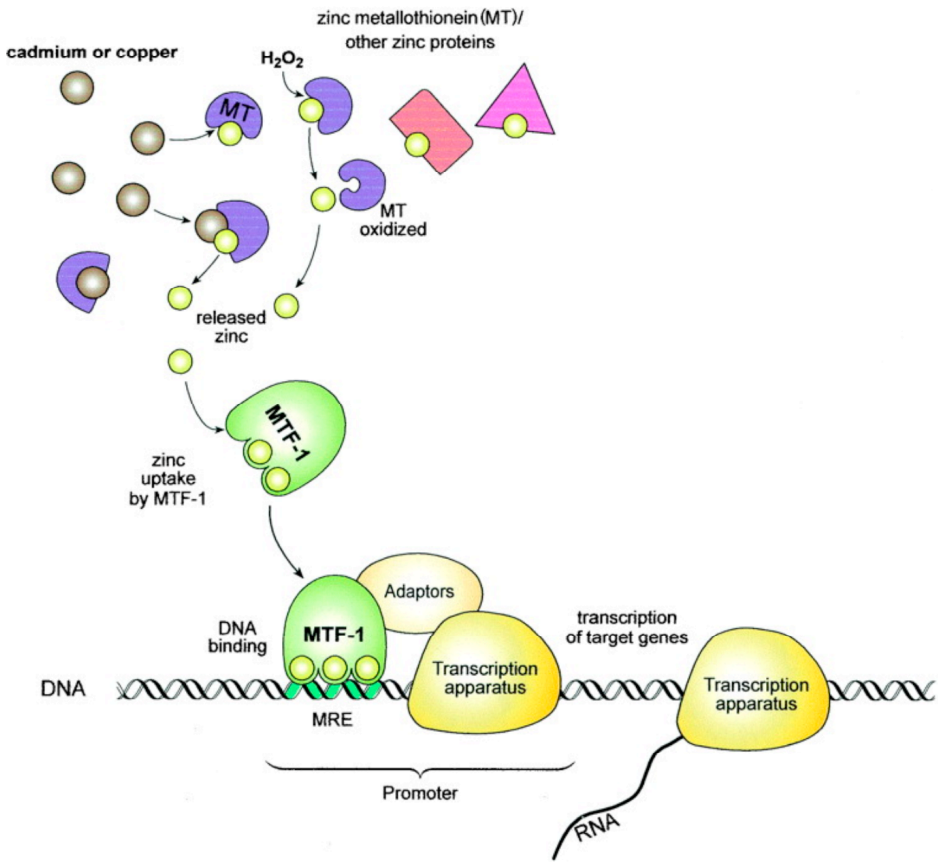


Figure 2. Model of MTF-1 activation derived from studies in a cell-free mammalian expression system.

Whereas zinc directly activates MTF-1, cadmium, copper, and hydrogen peroxide (H_2O_2) indirectly activate MTF-1 via zinc-loaded metallothionein (Zn-MT). Cadmium and copper bind to metallothionein (MT) with a much higher affinity than zinc, but due to the great abundance of the latter, the majority of metallothionein under physiological conditions is present as Zn-MT. H_2O_2 also induces zinc release from metallothionein via oxidation of sulfhydryl groups.
(Adapted from [Zhang, 2003])

MREs represent short DNA sequences in the enhancer/promoter regions of metallothionein genes and other target genes of MTF-1. MREs consist of the conserved core consensus sequence, TGCRCNC (R = A or G, and N = any nucleotide), followed by less conserved GC-rich sequences [Carter, 1984; Stuart, 1984; Stuart, 1985; Culotta, 1989]. A synthetic promoter composed of several MREs is sufficient for MTF-1-driven, metal responsible expression of target genes [Searle, 1985]. Mutations of single nucleotides in an MRE sequence can reduce or abolish its function [Searle, 1987; Radtke, 1993].

MTs are cysteine-rich, low molecular weight proteins that are able to bind a wide range of metals, thus playing protective roles but also serving as storage molecules [Margoshes, 1957; Stillman, 1995; Palmiter, 1998; Simpkins, 2000; Sigel, 2009; Blindauer, 2010]. The many cysteines which account for up to 30% of all amino acids, are arranged in characteristic CxC and CxxC motifs. MTs bind metals through the thiol group of their cysteines. Most of the known MT structures contain few intrinsic structure elements, because an ordered structure depends on the presence of bound metal ions [Blindauer, 2005].

MTs or MT-like proteins are present in a variety of phyla, from fungi to humans.

There are at least 12 MT genes in humans (*MT-I (A, B, E, F, G, H, M, L, X), MT-II, III* and *IV*). There is a growing evidence for human MTs being associated with various physiological conditions such as inflammation, cancer, diabetes and aging/longevity. The mouse has 4 MTs (MTI-IV). The expression of most metallothionein genes is regulated at transcriptional level in response to diverse cellular stresses [Kägi, 1991; Radtke, 1993; Palmiter 1998]. Mammalian MTs are predominantly zinc-thioneins, but copper and cadmium can easily replace zinc due to their higher affinity to MTs [Shaw, 1991]. Yeast has two metallothioneins [Winge, 1985; Hamer, 1985; Culotta, 1994]: the copper thionein CUP1 [Thiele, 1992], and CRS5, which shows dual metal-binding behaviour that is reminiscent of Zn-thioneins [Pagani, 2007]. Recently the classification for zinc- and copper-thioneins has been revised: now a stepwise gradation between extreme Zn-thioneins and Cu-thioneins has been proposed [Capdevila 2010]. *Drosophila* was previously shown to have four MTs, termed MtnA-D. Recently we characterized the fifth member of the *Drosophila* MT family, MtnE. MtnA-D are expressed mainly in the digestive tract and malpighian tubules, and their expression in the digestive tract is strongly boosted when larvae are fed with copper- or cadmium-supplemented food [Egli, 2006]. In comparison to wildtype *Drosophila,* a knockout mutant for all four *MtnA-D* genes is highly sensitive to copper and cadmium, to a lesser extent to zinc, but not to mercury or silver load or to copper depletion [Egli, 2006]. MREs are found in the promoter/enhancer regions of all the *Drosophila* MT genes and their expression is controlled by MTF-1.

Mammalian MTF-1 has several known interacting partners to regulate the expression of target genes in a positive or negative manner. Optimal expression of the mouse *MT-I* gene involves interactions of MTF-1 and the upstream stimulatory factor-1 (USF1) which binds to an E-box1 sequence at the *MT-I* promoter [Andrews, 2001b]. Hypoxia-inducible transcription factor-1alpha (HIF-1α) has been shown to be recruited, together with MTF-1, to the mouse *MT-I* promoter in response to hypoxia. HIF-1α functions as a coactivator of *MT-I* gene transcription by interacting with MTF-1 during hypoxia [Murphy, 1999; Murphy, 2008]. Hypoxia is also a stimulus for the nuclear translocation and interaction of MTF-1 with the p65 subunit of nuclear factor kappa B (NF-kB) on the promoter of placental growth factor (*PlGF*) [Cramer, 2005]. Mouse nuclear factor-I (NF-I/ CTF-I) is a repressor of MT-I expression via direct or indirect interaction with MTF-1 [Majumder, 2001; Wang, 2004].

Sp1, a general activator of transcription, can also bind GC-rich motifs overlapping with MRE sequences, possibly in a negative regulatory manner in competition with MTF-1 [Ogra, 2001]. Recently, a role for Sp1 and MTF-1 was postulated for copper-dependent regulation of the human *PRNP* gene (encoding prion protein which is involved in Creutzfeldt-Jakob disease, Gerstmann-Sträussler-Scheinker syndrome and fatal familial insomnia) [Bellingham, 2009]. The human *PRNP* promoter region contains a number of putative transcription factor-binding sites, including those for activators Sp1, AP1, AP2, and a CCAAT box binding [Mahal, 2001].

Upon zinc exposure a multiprotein complex containing MTF-1, the histone acetyltransferase p300/CBP (CREB-binding protein), and the transcription factor Sp1, is formed for transactivation of *MT-I* gene expression [Li, 2008]. MTF-1 is acetylated by p300; *in vitro* experiments showed that MTF-1 interacts with the CH3 and CH1 domains of p300 (V. Günther, W. Schaffner, unpublished data).

A global interaction analysis of *Drosophila* proteins in the yeast two-hybrid system [Giot, 2003] revealed three candidate interaction partners. Two of these, encoded by *CG6444* and *CG11591* [www.flybase.org] are closely related. Both proteins belong to the dumpy-30 (Dpy-30) family of proteins, first described in the nematode *C. elegans*, where it is an essential component of the *C. elegans* dosage compensation machinery

[Hsu, 1994; Hsu, 1995]. The human ortholog of Dpy-30 is a component of the human MLL2 (also termed ALR complex) [Issaeva, 2007].

We followed this up and substantiated the Dpy-30 homologs as interacting partners of MTF-1 as shown by [Vardanyan, 2008]. My contribution to this study includes dissections for the analysis of sperm motility, PCR analysis of *dpy30-L1* and *dpy30-L2* knockout fly strains and general manuscript handling.

References

1. **Andrews, G.** 2001a. Cellular zinc sensors: MTF-1 regulation of gene expression. Biometals **14:**223-37.

2. **Andrews, G., D. Lee, R. Ravindra, P. Lichtlen, M. Sirito, M. Sawadogo, and W. Schaffner.** 2001b. The transcription factors MTF-1 and USF1 cooperate to regulate mouse metallothionein-I expression in response to the essential metal zinc in visceral endoderm cells during early development. EMBO J **20:**1114-22.

3. **Auf der Maur, A., T. Belser, G. Elgar, O. Georgiev, and W. Schaffner.** 1999. Characterization of the transcription factor MTF-1 from the Japanese pufferfish (Fugu rubripes) reveals evolutionary conservation of heavy metal stress response. Biol Chem **380:**175-85.

4. **Beaudoin, J., and S. Labbé.** 2001. The fission yeast copper-sensing transcription factor Cuf1 regulates the copper transporter gene expression through an Ace1/Amt1-like recognition sequence. J Biol Chem **276:**15472-80.

5. **Beaudoin, J., A. Mercier, R. Langlois, and S. Labbé.** 2003. The Schizosaccharomyces pombe Cuf1 is composed of functional modules from two distinct classes of copper metalloregulatory transcription factors. J Biol Chem **278:**14565-77.

6. **Bellingham, S., L. Coleman, C. Masters, J. Camakaris, and A. Hill.** 2009. Regulation of prion gene expression by transcription factors SP1 and metal transcription factor-1. J Biol Chem **284:**1291-301.

7. **Blaiseau, P., E. Lesuisse, and J. Camadro.** 2001. Aft2p, a novel iron-regulated transcription activator that modulates, with Aft1p, intracellular iron use and resistance to oxidative stress in yeast. J Biol Chem **276:**34221-6.

8. **Blindauer, C., and O. Leszczyszyn.** 2010. Metallothioneins: unparalleled diversity in structures and functions for metal ion homeostasis and more. Nat Prod Rep **27:**720-41.

9. **Blindauer, C., and P. Sadler.** 2005. How to hide zinc in a small protein. Acc Chem Res **38:**62-9.

10. **Brugnera, E., O. Georgiev, F. Radtke, R. Heuchel, E. Baker, G. Sutherland, and W. Schaffner.** 1994. Cloning, chromosomal mapping and characterization of the human metal-regulatory transcription factor MTF-1. Nucleic Acids Res **22:**3167-73.

11. **Buchman, C., P. Skroch, J. Welch, S. Fogel, and M. Karin.** 1989. The CUP2 gene product, regulator of yeast metallothionein expression, is a copper-activated DNA-binding protein. Mol Cell Biol **9:**4091-5.

12. **Capdevila, M., O. Palacios, and S. Atrian.** 2010. The zn- or cu-thionein character of a metallothionein determines its metal load when synthesized in physiological (metal-unsupplemented) conditions. Bioinorg Chem Appl**:**541829.

13. **Carter, A., B. Felber, M. Walling, M. Jubier, C. Schmidt, and D. Hamer.** 1984. Duplicated heavy metal control sequences of the mouse metallothionein-I gene. Proc Natl Acad Sci U S A **81:**7392-6.

14. **Chen, X., B. Zhang, P. Harmon, W. Schaffner, D. Peterson, and D. Giedroc.** 2004. A novel cysteine cluster in human metal-responsive transcription factor 1 is required for heavy metal-induced transcriptional activation in vivo. J Biol Chem **279:**4515-22.

15. **Courel, M., S. Lallet, J. Camadro, and P. Blaiseau.** 2005. Direct activation of genes involved in intracellular iron use by the yeast iron-responsive transcription factor Aft2 without its paralog Aft1. Mol Cell Biol **25:**6760-71.

16. **Cramer, M., I. Nagy, B. Murphy, M. Gassmann, M. Hottiger, O. Georgiev, and W. Schaffner.** 2005. NF-kappaB contributes to transcription of placenta growth factor and interacts with metal responsive transcription factor-1 in hypoxic human cells. Biol Chem **386:**865-72.

17. **Culotta, V., and D. Hamer.** 1989. Fine mapping of a mouse metallothionein gene metal response element. Mol Cell Biol **9:**1376-80.

18. **Culotta, V., W. Howard, and X. Liu.** 1994. CRS5 encodes a metallothionein-like protein in Saccharomyces cerevisiae. J Biol Chem **269:**25295-302.

19. **Das, S., and K. Ray.** 2006. Wilson's disease: an update. Nat Clin Pract Neurol **2:**482-93.

20. **de Bie, P., P. Muller, C. Wijmenga, and L. Klomp.** 2007. Molecular pathogenesis of Wilson and Menkes disease: correlation of mutations with molecular defects and disease phenotypes. J Med Genet **44:**673-88.

21. **Dufner-Beattie, J., F. Wang, Y. Kuo, J. Gitschier, D. Eide, and G. Andrews.** 2003. The acrodermatitis enteropathica gene ZIP4 encodes a tissue-specific, zinc-regulated zinc transporter in mice. J Biol Chem **278:**33474-81.

22. **Egli, D., A. Selvaraj, H. Yepiskoposyan, B. Zhang, E. Hafen, O. Georgiev, and W. Schaffner.** 2003. Knockout of 'metal-responsive transcription factor' MTF-1 in Drosophila by homologous recombination reveals its central role in heavy metal homeostasis. EMBO J **22:**100-8.

23. **Egli, D., H. Yepiskoposyan, A. Selvaraj, K. Balamurugan, R. Rajaram, A. Simons, G. Multhaup, S. Mettler, A. Vardanyan, O. Georgiev, and W. Schaffner.** 2006. A family knockout of all four Drosophila metallothioneins reveals a central role in copper homeostasis and detoxification. Mol Cell Biol **26:**2286-96.

24. **Giedroc, D., X. Chen, and J. Apuy.** 2001. Metal response element (MRE)-binding transcription factor-1 (MTF-1): structure, function, and regulation. Antioxid Redox Signal **3:**577-96.

25. **Giot, L., J. Bader, C. Brouwer, A. Chaudhuri, B. Kuang, Y. Li, Y. Hao, C. Ooi, B. Godwin, E. Vitols, G. Vijayadamodar, P. Pochart, H. Machineni, M. Welsh, Y. Kong, B. Zerhusen, R. Malcolm, Z. Varrone, A. Collis, M. Minto, S. Burgess, L. McDaniel, E. Stimpson, F. Spriggs, J. Williams, K. Neurath, N. Ioime, M. Agee, E. Voss, K. Furtak, R. Renzulli, N. Aanensen, S. Carrolla, E. Bickelhaupt, Y. Lazovatsky, A. DaSilva, J. Zhong, C. Stanyon, R. J. Finley, K. White, M. Braverman, T. Jarvie, S. Gold, M. Leach, J. Knight, R. Shimkets, M. McKenna, J. Chant, and J. Rothberg.** 2003. A protein interaction map of Drosophila melanogaster. Science **302:**1727-36.

26. **Hamer, D., D. Thiele, and J. Lemontt.** 1985. Function and autoregulation of yeast copperthionein. Science **228:**685-90.

27. **Hamza, I., J. Prohaska, and J. Gitlin.** 2003. Essential role for Atox1 in the copper-mediated intracellular trafficking of the Menkes ATPase. Proc Natl Acad Sci U S A **100:**1215-20.

28. **Heuchel, R., F. Radtke, O. Georgiev, G. Stark, M. Aguet, and W. Schaffner.** 1994. The transcription factor MTF-1 is essential for basal and heavy metal-induced metallothionein gene expression. EMBO J **13:**2870-5.

29. **Hsu, D., P. Chuang, and B. Meyer.** 1995. DPY-30, a nuclear protein essential early in embryogenesis for Caenorhabditis elegans dosage compensation. Development **121:**3323-34.

30. **Hsu, D., and B. Meyer.** 1994. The dpy-30 gene encodes an essential component of the Caenorhabditis elegans dosage compensation machinery. Genetics **137:**999-1018.

31. **Issaeva, I., Y. Zonis, T. Rozovskaia, K. Orlovsky, C. Croce, T. Nakamura, A. Mazo, L. Eisenbach, and E. Canaani.** 2007. Knockdown of ALR (MLL2) reveals ALR target genes and leads to alterations in cell adhesion and growth. Mol Cell Biol **27:**1889-903.

32. **Jungmann, J., H. Reins, J. Lee, A. Romeo, R. Hassett, D. Kosman, and S. Jentsch.** 1993. MAC1, a nuclear regulatory protein related to Cu-dependent transcription factors is involved in Cu/Fe utilization and stress resistance in yeast. EMBO J **12:**5051-6.

33. **Kägi, J.** 1991. Overview of metallothionein. Methods Enzymol **205:**613-26.

34. **Li, Y., T. Kimura, R. Huyck, J. Laity, and G. Andrews.** 2008. Zinc-induced formation of a coactivator complex containing the zinc-sensing transcription factor MTF-1, p300/CBP, and Sp1. Mol Cell Biol **28:**4275-84.

35. **Lichtlen, P., and W. Schaffner.** 2001. Putting its fingers on stressful situations: the heavy metal-regulatory transcription factor MTF-1. Bioessays **23:**1010-7.

36. **Lindert, U., M. Cramer, M. Meuli, O. Georgiev, and W. Schaffner.** 2009. Metal-responsive transcription factor 1 (MTF-1) activity is regulated by a nonconventional nuclear localization signal and a metal-responsive transactivation domain. Mol Cell Biol **29:**6283-93.

37. **Lyons, T., A. Gasch, L. Gaither, D. Botstein, P. Brown, and D. Eide.** 2000. Genome-wide characterization of the Zap1p zinc-responsive regulon in yeast. Proc Natl Acad Sci USA **97:**7957-62.

38. **Mahal, S., E. Asante, M. Antoniou, and J. Collinge.** 2001. Isolation and functional characterisation of the promoter region of the human prion protein gene. Gene **268:**105-14.

39. **Majumder, S., K. Ghoshal, R. Gronostajski, and S. Jacob.** 2001. Downregulation of constitutive and heavy metal-induced metallothionein-I expression by nuclear factor I. Gene Expr **9:**203-15.

40. **Margoshes M., and V. B. L.** 1957. A cadmium protein from equine kidney cortex, p. 4813-4814, J. Am. Chem. Soc.

41. **Menkes, J., M. Alter, G. Steigleder, D. Weakley, and J. Sung.** 1962. A sex-linked recessive disorder with retardation of growth, peculiar hair, and focal cerebral and cerebellar degeneration. Pediatrics **29:**764-79.

42. **Murphy, B., G. Andrews, D. Bittel, D. Discher, J. McCue, C. Green, M. Yanovsky, A. Giaccia, R. Sutherland, K. Laderoute, and K. Webster.** 1999. Activation of metallothionein gene expression by hypoxia involves metal response elements and metal transcription factor-1. Cancer Res **59:**1315-22.

43. **Murphy, B., T. Kimura, B. Sato, Y. Shi, and G. Andrews.** 2008. Metallothionein induction by hypoxia involves cooperative interactions between metal-responsive transcription factor-1 and hypoxia-inducible transcription factor-1alpha. Mol Cancer Res **6:**483-90.

44. **Ogra, Y., K. Suzuki, P. Gong, F. Otsuka, and S. Koizumi.** 2001. Negative regulatory role of Sp1 in metal responsive element-mediated transcriptional activation. J Biol Chem **276:**16534-9.

45. **Pagani, A., L. Villarreal, M. Capdevila, and S. Atrian.** 2007. The Saccharomyces cerevisiae Crs5 Metallothionein metal-binding abilities and its role in the response to zinc overload. Mol Microbiol **63:**256-69.

46. **Palmiter, R.** 1998. The elusive function of metallothioneins. Proc Natl Acad Sci U S A **95:**8428-30.

47. **Pelletier, B., J. Beaudoin, Y. Mukai, and S. Labbé.** 2002. Fep1, an iron sensor regulating iron transporter gene expression in Schizosaccharomyces pombe. J Biol Chem **277:**22950-8.

48. **Pelletier, B., A. Trott, K. Morano, and S. Labbé.** 2005. Functional characterization of the iron-regulatory transcription factor Fep1 from Schizosaccharomyces pombe. J Biol Chem **280:**25146-61.

49. **Radtke, F., O. Georgiev, H. Müller, E. Brugnera, and W. Schaffner.** 1995. Functional domains of the heavy metal-responsive transcription regulator MTF-1. Nucleic Acids Res **23:**2277-86.

50. **Radtke, F., R. Heuchel, O. Georgiev, M. Hergersberg, M. Gariglio, Z. Dembic, and W. Schaffner.** 1993. Cloned transcription factor MTF-1 activates the mouse metallothionein I promoter. EMBO J **12:**1355-62.

51. **Rivera-Mancía, S., I. Pérez-Neri, C. Ríos, L. Tristán-López, L. Rivera-Espinosa, and S. Montes.** 2010. The transition metals copper and iron in neurodegenerative diseases. Chem Biol Interact **186:**184-99.

52. **Robinson, N., and D. Winge.** 2010. Copper metallochaperones. Annu Rev Biochem **79:**537-62.

53. **Rosenzweig, A.** 2002. Metallochaperones: bind and deliver. Chem Biol **9:**673-7.

54. **Schauer, K., C. Muller, M. Carrière, A. Labigne, C. Cavazza, and H. De Reuse.** 2010. The Helicobacter pylori GroES cochaperonin HspA functions as a specialized nickel chaperone and sequestration protein through its unique C-terminal extension. J Bacteriol **192:**1231-7.

55. **Searle, P., G. Stuart, and R. Palmiter.** 1985. Building a metal-responsive promoter with synthetic regulatory elements. Mol Cell Biol **5:**1480-9.

56. **Searle, P., G. Stuart, and R. Palmiter.** 1987. Metal regulatory elements of the mouse metallothionein-I gene. Experientia Suppl **52:**407-14.

57. **Shakoury-Elizeh, M., J. Tiedeman, J. Rashford, T. Ferea, J. Demeter, E. Garcia, R. Rolfes, P. Brown, D. Botstein, and C. Philpott.** 2004. Transcriptional remodeling in response to iron deprivation in Saccharomyces cerevisiae. Mol Biol Cell **15:**1233-43.

58. **Shaw, C. r., M. Savas, and D. Petering.** 1991. Ligand substitution and sulfhydryl reactivity of metallothionein. Methods Enzymol **205:**401-14.

59. **Sigel A., Sigel H., and Sigel R.** 2009. Metal Ions in Life Sciences. RSC publishing.

60. **Simpkins, C.** 2000. Metallothionein in human disease. Cell Mol Biol (Noisy-le-grand) **46:**465-88.

61. **Stillman, M. J.** 1995. Metallothioneins, p. 461-511, **Coordination Chemistry Reviews**, vol. 144.

62. **Stola, M., F. Musiani, S. Mangani, P. Turano, N. Safarov, B. Zambelli, and S. Ciurli.** 2006. The nickel site of Bacillus pasteurii UreE, a urease metallo-chaperone, as revealed by metal-binding studies and X-ray absorption spectroscopy. Biochemistry **45:**6495-509.

63. **Stuart, G., P. Searle, H. Chen, R. Brinster, and R. Palmiter.** 1984. A 12-base-pair DNA motif that is repeated several times in metallothionein gene promoters confers metal regulation to a heterologous gene. Proc Natl Acad Sci U S A **81:**7318-22.

64. **Stuart, G., P. Searle, and R. Palmiter.** 1985. Identification of multiple metal regulatory elements in mouse metallothionein-I promoter by assaying synthetic sequences. Nature **317:**828-31.

65. **Thiele, D.** 1992. Metal-regulated transcription in eukaryotes. Nucleic Acids Res **20:**1183-91.

66. **Tümer, Z., and L. Møller.** 2010. Menkes disease. Eur J Hum Genet **18:**511-8.

67. **Vardanyan, A., L. Atanesyan, D. Egli, S. Raja, M. Steinmann-Zwicky, R. Renkawitz-Pohl, O. Georgiev, and W. Schaffner.** 2008. Dumpy-30 family members as determinants of male fertility and interaction partners of metal-responsive transcription factor 1 (MTF-1) in Drosophila. BMC Dev Biol **8:**68.

68. **Walker, J., R. Tsivkovskii, and S. Lutsenko.** 2002. Metallochaperone Atox1 transfers copper to the NH2-terminal domain of the Wilson's disease protein and regulates its catalytic activity. J Biol Chem **277:**27953-9.

69. **Wang, Y., I. Lorenzi, O. Georgiev, and W. Schaffner.** 2004. Metal-responsive transcription factor-1 (MTF-1) selects different types of metal response elements at low vs. high zinc concentration. Biol Chem **385:**623-32.

70. **Westin, G., and W. Schaffner.** 1988. A zinc-responsive factor interacts with a metal-regulated enhancer element (MRE) of the mouse metallothionein-I gene. EMBO J **7:**3763-70.

71. **Wilson S.** 1912. Progressive lenticular degeneration: a familial nervous disease associated with cirrhosis of the liver., p. 295–507, vol. *3***4**, Brain.

72. **Winge, D., K. Nielson, W. Gray, and D. Hamer.** 1985. Yeast metallothionein. Sequence and metal-binding properties. J Biol Chem **260:**14464-70.

73. **Yamaguchi-Iwai, Y., A. Dancis, and R. Klausner.** 1995. AFT1: a mediator of iron regulated transcriptional control in Saccharomyces cerevisiae. EMBO J **14:**1231-9.

74. **Yamaguchi-Iwai, Y., M. Serpe, D. Haile, W. Yang, D. Kosman, R. Klausner, and A. Dancis.** 1997. Homeostatic regulation of copper uptake in yeast via direct binding of MAC1 protein to upstream regulatory sequences of FRE1 and CTR1. J Biol Chem **272:**17711-8.

75. **Yamaguchi-Iwai, Y., R. Stearman, A. Dancis, and R. Klausner.** 1996. Iron-regulated DNA binding by the AFT1 protein controls the iron regulon in yeast. EMBO J **15:**3377-84.

76. **Yepiskoposyan, H., D. Egli, T. Fergestad, A. Selvaraj, C. Treiber, G. Multhaup, O. Georgiev, and W. Schaffner.** 2006. Transcriptome response to heavy metal stress in Drosophila reveals a new zinc transporter that confers resistance to zinc. Nucleic Acids Res **34:**4866-77.

77. **Zhang, B., D. Egli, O. Georgiev, and W. Schaffner.** 2001. The Drosophila homolog of mammalian zinc finger factor MTF-1 activates transcription in response to heavy metals. Mol Cell Biol **21:**4505-14.

78. **Zhang, B., O. Georgiev, M. Hagmann, C. Günes, M. Cramer, P. Faller, M. Vasák, and W. Schaffner.** 2003. Activity of metal-responsive transcription factor 1 by toxic heavy metals and H2O2 in vitro is modulated by metallothionein. Mol Cell Biol **23:**8471-85.

79. **Zhao, H., and D. Eide.** 1997. Zap1p, a metalloregulatory protein involved in zinc-responsive transcriptional regulation in Saccharomyces cerevisiae. Mol Cell Biol **17:**5044-52.

Title:

Characterization of MtnE, the fifth metallothionein member in *Drosophila*

Authors: Lilit Atanesyan[1, 2], Oleg Georgiev[1], Walter Schaffner[1]

[1]Institute of Molecular Life Sciences, University of Zurich,
Winterthurerstrasse 190, CH-8057, Zurich, Switzerland

[2]Life Science Zurich Graduate School, University of Zurich/ETH Zurich, Switzerland

To whom correspondence should be addressed:

Walter Schaffner
Institute of Molecular Life Sciences, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland;

phone: +41 635 31 50/51; fax: +41 44 635 68 11; e-mail: walter.schaffner@imls.uzh.ch

Abstract

Metallothioneins (MTs) constitute a family of cysteine-rich, low molecular weight metal-binding proteins. These proteins occur in almost all forms of life, they bind physiological metals such as zinc and copper, but also non-essential, toxic heavy metals such as cadmium, mercury and silver. MT expression is regulated at the transcriptional level by the metal-responsive transcription factor-1 (MTF-1), which binds to the metal response elements (MREs) at the enhancer/promoter regions of MT genes. *Drosophila* was shown before to have four MTs, namely MtnA, MtnB, MtnC and MtnD. Here we characterize a newly found fifth member of *Drosophila* MT gene family, coding for metallothionein E (MtnE). The *MtnE* transcription unit is located head to head with the one of *MtnD*. The intervening sequence contains four MREs and mutations of individual MREs have different effects on the expression of these genes.

Introduction

Metallothioneins (MTs) are cysteine-rich (up to 30%), low molecular weight proteins that are able to bind a wide range of metals including cadmium, zinc, mercury, copper and silver [Stillman, 1995; Palmiter, 1998; Simpkins, 2000; Sigel, 2009; Blindauer, 2010]. Metallothioneins were first characterized in 1957 by Vallee and Margoshes as cadmium-binding proteins in the horse renal cortex [Margoshe, 1957]. MTs play an essential role in heavy metal detoxification and the regulation of the metabolism of essential trace metals; they are also involved in the protection against free radicals and oxidative stress [Sato, 1993; Sato, 2010; Chivaverini, 2010]. MTs are found in all eukaryotes and also in some prokaryotes [Robinson, 2001]. Humans have at least one dozen MT genes, while the mouse has four of them. *Drosophila* was previously shown to have four MTs (MtnA-D) expressed mainly in the digestive tract [Egli, 2006]. *S. cerevisiae* has two MTs, CUP1 and CRS5 [Winge, 1985; Hamer, 1985].

Metal-responsive transcription factor MTF-1 is a Cys2-His2 zinc-finger protein that plays a central role in the heavy metal-induced transcriptional response from insects to humans [Westin, 1988; Radtke, 1993; Auf der Maur, 1999; Andrews, 2001; Lichtlen, 2001; Giedroc, 2001]. MTF-1 is an essential, ubiquitously expressed protein in mammals; *MTF-1* null mutant mouse embryos develop severe liver degeneration and die *in utero* at approximately day 14 of gestation [Günes, 1998]. *MTF-1* knockout *Drosophila* are viable, and fertile under standard laboratory conditions, but sensitive to heavy metal load and to copper starvation [Egli, 2003].

The expression of metallothioneins is regulated at the transcriptional level by MTF-1 [Westin, 1988; Radtke, 1993; Heuchel, 1994]. Upon metal stress, MTF-1 binds to short DNA sequences in the enhancer/promoter regions of metallothionein genes and other target genes, termed metal-response elements (MREs). MREs consist of the highly conserved core consensus sequence, TGCRCNC (R = A or G, and N = any nucleotide), followed by less conserved GC-rich sequences [Carter, 1984; Stuart, 1984; Stuart, 1985; Culotta, 1989; Wang, 2004]. A synthetic promoter composed of several MRE sequences is sufficient for MTF-1-driven expression of target genes upon metal induction [Searle, 1985]. Mutations of single nucleotides in MRE sequences can reduce or abolish its

function [Searle, 1987; Radtke, 1993]. MTF-1 plays a role not only in heavy metal stress but also in other cell stress conditions such as oxidative stress and hypoxia [Murphy, 1999; Dalton, 2000].

The existence of a fifth *Drosophila* metallothionein gene, annotated as *CG42872* in the flybase, was recently brought to our attention (S. Celniker, personal communication; www.flybase.org). *CG42872* is highly similar to *Drosophila* metallothionein genes (*MtnA-D*) [Egli, 2003; Egli, 2006] and therefore it was named metallothionein E (*MtnE*) (Figure 1). We cloned the genomic region containing the *MtnE* and *MtnD* genes, which are adjacent to each other. In order to study the expression profiles, *MtnD* and *MtnE* were differentially tagged with GFP and/or mCherry fluorescent tags.

Materials and methods

Annotation of MT genes in insects

MT-type genes in the different insect genomes were identified by standard BLAST searches (NCBI BLAST and Flybase BLAST) against the sequenced genomes and against protein databases. Sequence alignments were generated with the CLC Sequence Viewer, Version 6.3, with standard parameters.

Fly culture

1 liter of fly food was composed of 55 g cornmeal (Maisgriess 54.401.025, Meyerhans Hotz AG), 100 g yeast (Hefe Schweiz AG), 75 g sugar (Dextrose monohydrate), 8 g agar (Insectagar type ZN5) and 15 ml anti-fungal Nipagin (nipagin 33 g/l, nipasol 66 g/l in 96% ethanol). For experiments, food was supplemented with $CuSO_4$, $HgCl_2$ or $ZnSO_4$ to the concentrations indicated on the figures. Flies were raised at 25ºC and 65% humidity.

Cell culture and transfection

Drosophila Schneider 2 (S2) cells were cultured in Schneider's *Drosophila* medium, containing 10% heat-inactivated fetal bovine serum (Biochrom AG), 100 U/ml penicillin, 100 μg/ml streptomycin and 2 mM L-glutamine (Invitrogen). Cells of an exponentially growing 10 cm culture Petri dish were transfected with 10 μg of the particular expression clone and 2.5 μg of reference plasmid. The calcium-phosphate co-precipitation method was used for transfection. Cells were harvested for RNA extraction 48 hours post transfection. For the induction of MT expression, $CuSO_4$ was added to the cell incubation media to 100 μM final concentration 24 hours post transfection.

S1 nuclease protection assay

Isolation of RNA and the S1 nuclease protection assay were done as described previously [Weaver, 1979; Westin, 1987]. Signals were visualized using the fluorescent image analyzer FLA-7000 and quantified using the ImageGauge software (Fujifilm Life Science). Reporter signals were normalized to the reference signals.

Tissue preparation and microscopy

To harvest guts, wandering 3rd instar larvae were dissected in ice-cold PBS, pH 7.4. Tissues were directly mounted in *Drosophila* Ringer's solution (3 mM $CaCl_2$, 46 mM NaCl, 182 mM KCl, 10 mM Tris, pH 7.2) and microphotographed (Figure 6). FITC and TRITC filter epifluorescence pictures from larval guts were taken using a Zeiss Axioplan 2 microscope and a Zeiss Axiocam MRm camera with 5x magnification. For the whole-larvae imaging, wandering 3rd instar larvae were placed on glass slide and immobilized by incubating them shortly (~1 min) at 4ºC. The pictures were taken using a Leica DMRB microscope and a Zeiss Axiocam color camera with 2.5x magnification.

Plasmid construction and transgenic fly strains

Transgenic fly lines were generated using ϕC31-mediated transgenesis [Bischof, 2001; Fish, 2007]. In all experiments the *86Fb* (chromosome 3R) landing site was used for plasmid integration. All the flies carrying the transgene integration site also expressed the ϕC31 integrase, which was removed by crossing out after the integration event occurred. To generate *MtnD* and *MtnE* transgenes, a DNA segment containing the two transcription units and flanking sequences corresponding to the genomic region of *D. melanogaster* (coordinates 3R: complement (16360941..16363020)) was cloned into a vector containing an *attB* site for integration into the specific genomic locus, and a *miniwhite* marker gene. Plasmid sequences and detailed cloning strategies are available upon request.

For the overexpression of dMTF-1, we used the constitutively active *Drosophila* tubulin promoter (Tub). The Tub-dMTF-1-VSV transgene was integrated into the *51D* locus on the 2nd chromosome. The null allele for *dMTF-1* is termed *D4*.

Results and discussion

MtnE transcription is induced by heavy metals

Transgenic flies expressing the GFP- or mCherry-tagged *MtnE* gene on a *MTF-1* wild-type background showed stronger fluorescence in the presence of heavy metals, namely copper and cadmium (Figure 3). An elevated expression of *MtnE* was also observed when the larvae were raised on food supplemented with 1 mM zinc (data not shown). Iron, of all the tested metals, had at most a marginal effect on MtnE expression. The expression of MtnD was also metal responsive, whereby the *MtnE* transgene yielded an even stronger fluorescence than *MtnD* both on NF and Cu-containing food (Figure 4). Overexpression of *dMTF-1* in the heterozygous *dMTF-1* background is greatly inducing the expression of MtnE (Figure 5). From these results we conclude that the expression of MtnE, like the other members of *Drosophila* metallothionein family, is regulated by *Drosophila* metal-responsive transcription factor-1 (dMTF-1).

Expression pattern of *MtnE* and *MtnD* genes largely overlaps

We further investigated the expression of MtnE-GFP and MtnD-mCherry proteins in *Drosophila* larval tissues. Like that of other *Drosophila* metallothioneins MtnE expression was observed in the intestine of adult flies and larvae on normal food (Figure 6A) and was boosted by copper-containing food (Figure 3, Figure 6B).

The expression patterns of *MtnD* and *MtnE* genes mostly overlap in the *Drosophila* larval gut, but nevertheless display some differences. *MtnE* is more widely expressed, including foregut and posterior midgut, even on standard food. As indicated by blue arrows, copper induces the expression of *MtnD* in a sub-region of the middle midgut that contains “cuprophilic cells” or “copper cells”. Copper cells are known to accumulate metal ions following copper or cadmium intoxication [Filshie, 1971; Egli, 2006]. *MtnD* is also constantly expressed in the so-called iron cell region [Brenner, 1997], regardless of the presence of metals (Figure 6A, B).

Mutations of MREs alter the expression of MtnD and MtnE

As mentioned before, there are four MRE sequence motifs in the DNA segment between *MtnD* and *MtnE* transcription units, that are most probably shared functionally among these two genes. We mutated single MREs (Figure 7A) to elucidate their role in the metal-induced expression of MtnD and MtnE. Preliminary results indicate that individual MREs are preferentially (but not exclusively) used by one or the other gene (Figure 7B and data not shown).

Conclusions

The family of *Drosophila* metallothioneins is completed with a new member, MtnE. All five metallothionein genes in *Drosophila* are located on the same chromosome and they harbour many structural and regulatory similarities. We showed that MtnE, like four previously described metallothioneins, is expressed in response to copper and cadmium in the intestine. *MtnE* shares a cluster of metal response elements with the divergently transcribed *MtnD*, and similar to the other metallothioneins its expression depends on MTF-1. *MtnA-D* knockout (*qMtn**) flies are viable and fertile but sensitive to elevated concentrations of copper or cadmium, to a lesser extent to zinc, but not to mercury or silver load or copper depletion [Egli, 2006]. The new member MtnE might complement the function of the others by widening the spectrum of metals that can be detoxified.

Figure legends

Figure 1. *D. melanogaster* metallothioneins (MtnA-E).

A. Nucleotide sequences of protein-coding regions (CDS) of MTs, which are located on *D. melanogaster* chromosome 3R.
B. Note that MtnB, C, D and E sequences are closely related to each other and different from MtnA.

Figure 2. *MtnD* and *MtnE* gene region of *D. melanogaster* with regulatory metal-response elements (MREs).
The *MtnD* and *MtnE* genes are divergently transcribed. Four MREs are located between their transcription start sites. MREs are indicated in red and named MRE 1-4; nucleotides in capital letters represent the mRNA, blue letters indicate the protein-coding region.

Figure 3. *MtnE-GFP* and *MtnE-mCherry* transgenes are expressed mostly in the larval digestive tract.
Transgenic flies were raised on normal food (NF), copper (Cu) - and cadmium (Cd) - containing food; the pictures were taken with 175 ms and 100 ms exposure time for GFP- and mCherry-tagged MTs, respectively. The concentration of metals in fly food is indicated on the top of each panel.

Figure 4. Expression levels are different for MtnD and MtnE.
MtnE is expressed more strongly than MtnD; in each box on the left side are *MtnD* transgenes tagged with mCherry or GFP, on the right are *MtnE-mCherry/GFP* transgenes. mCherry-tagged transgenes show strong expression which is even evident on bright field (left vertical panel). The expression levels are comparable within the same experiment (NF or Cu-food) but not necessarily between different experiments.

Figure 5. Expression of *MtnE* is modulated by dMTF-1.
The expression of the *MtnE-GFP* is stronger when dMTF-1 is overexpressed on the *dMTF-1* single knockout allele (*D4*) background in comparison to the control (no overexpression).

Figure 6. Expression pattern of metallothionein reporter constructs.

Larvae were raised on normal food (**A**) or 100 μM copper-containing food (**B**). The intensity of fluorescence is not comparable between (**A**) and (**B**), as well as for different fluorescent tags. Copper cells are indicated with blue arrows, orange arrows indicate the region of iron cells/midgut constriction.

Figure 7. Schematic view of the *MtnD* and *MtnE* genomic regions: the positions and orientations of the four MREs.

A. Wt and mut1-4 indicate the wildtype and mutated genomic regions, respectively.

B. Transcript levels of MRE mutants (mut1-4) of *MtnD-GFP* measured by S1 nuclease protection assay.

Figures

Figure 1.

A.

```
MtnA CDS  ATGCCTTGCC CA---TGCGG AAGCGGATGC AAATGCGCCA GCCAGGCCAC CAAGGGATCC 57
MtnB CDS  ATGGTTTGCA AGGGTTGTGG AACAAACTGC CAGTGCTCGG CCCAAAAGTG CGGGGACAAC 60
MtnC CDS  ATGGTTTGCA AAGGCTGCGG AACAAACTGC AAGTGCCAGG ACACCAAGTG CGGCGACAAT 60
MtnD CDS  ATGGGTTGCA AGGCTTGTGG AACAAACTGC CAGTGCTCCG CCACCAAGTG CGGTGACAAC 60
MtnE CDS  ATGCCTTGCA AGGGATGTGG AAACAACTGC CAGTGCTCAG CCGGAAAGTG CGGAGGTAAC 60

MtnA CDS  TGCAACTGCG GATCTGACTG CAAGTGCGGC GGCGACAAGA AATCCGCC-- ----TGCGGC 111
MtnB CDS  TGCGCCTGCA ACAAGGATTG CCAGTGCGTT TGCAAGAATG GGCCCAAGGA CCAGTGCTGC 120
MtnC CDS  TGCGCCTGTA ATCAGGACTG CAAGTGCGTG TGCAAGAATG GCCCCAAAGA TCAGTGTTGC 120
MtnD CDS  TGCGCCTGCA GCCAGCAGTG CCAGTGCTCC TGCTAGAACG GACCCAAGGA CAAGTGCTGC 120
MtnE CDS  TGCGCCGGAA ACAGCCAATG CCAATGCGCC GCCAAGACGG GAGCCAAG-- ----TGCTGC 114

MtnA CDS  TGCTCCGAG- --TGA 123
MtnB CDS  AGCAACAAA- --TAA 132
MtnC CDS  AAGAGCAAG- --TAG 132
MtnD CDS  TCCACCAAAA ACTAG 135
MtnE CDS  CAGGCCAAG- --TGA 126
```

B.

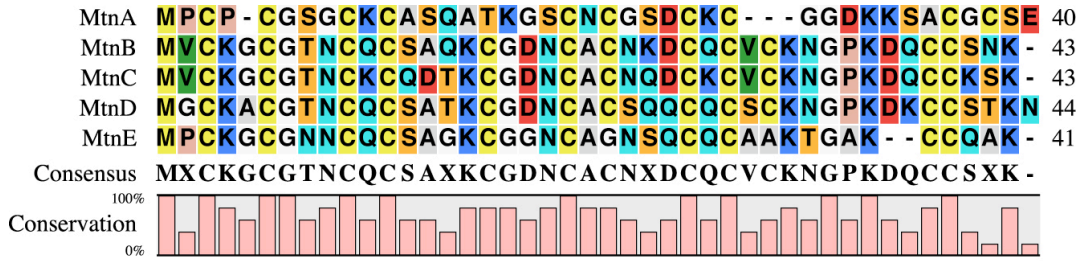


Figure 2.

Figure 3.

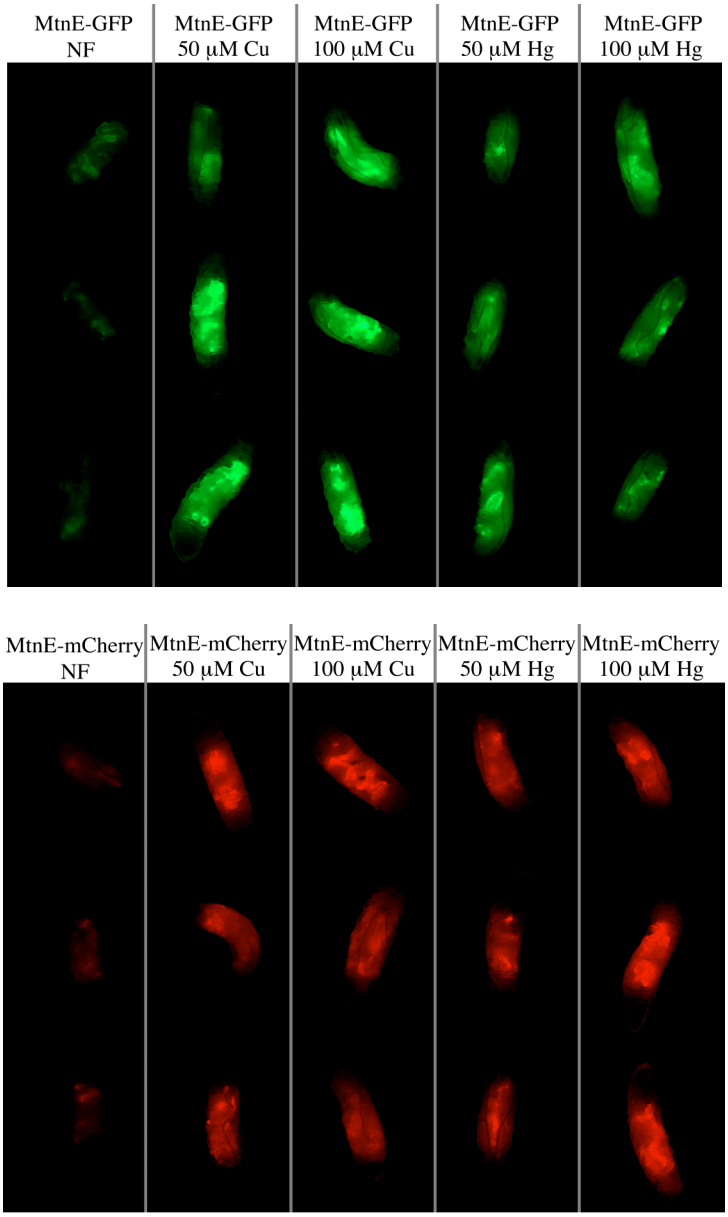


Figure 4.

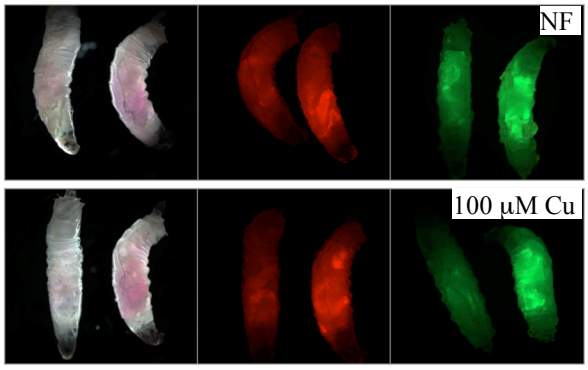


Figure 5.

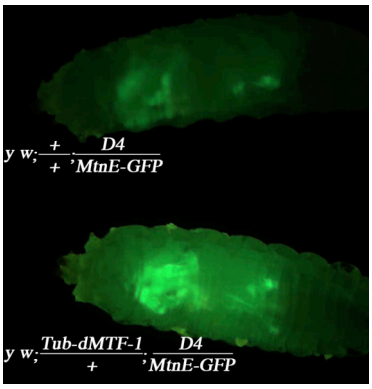
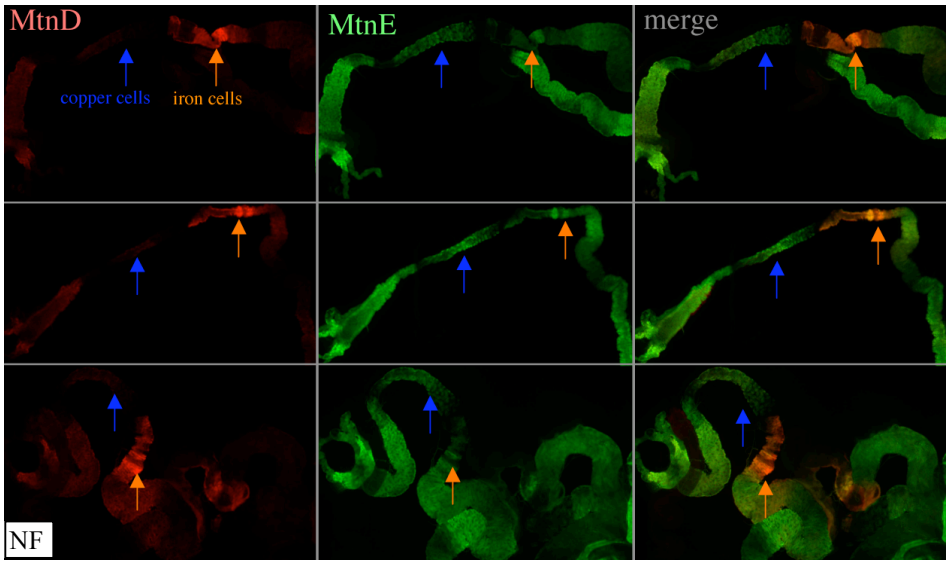


Figure 6.

A.



B.

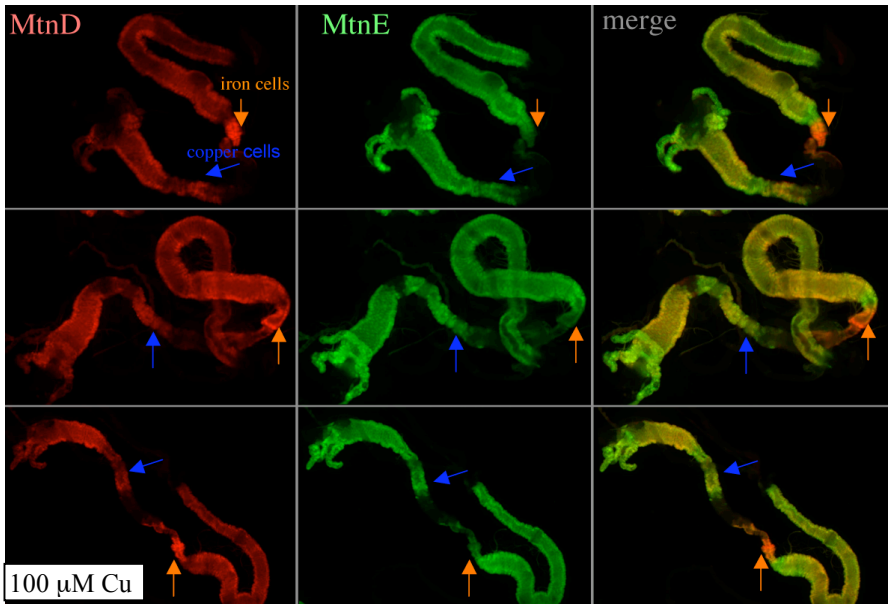
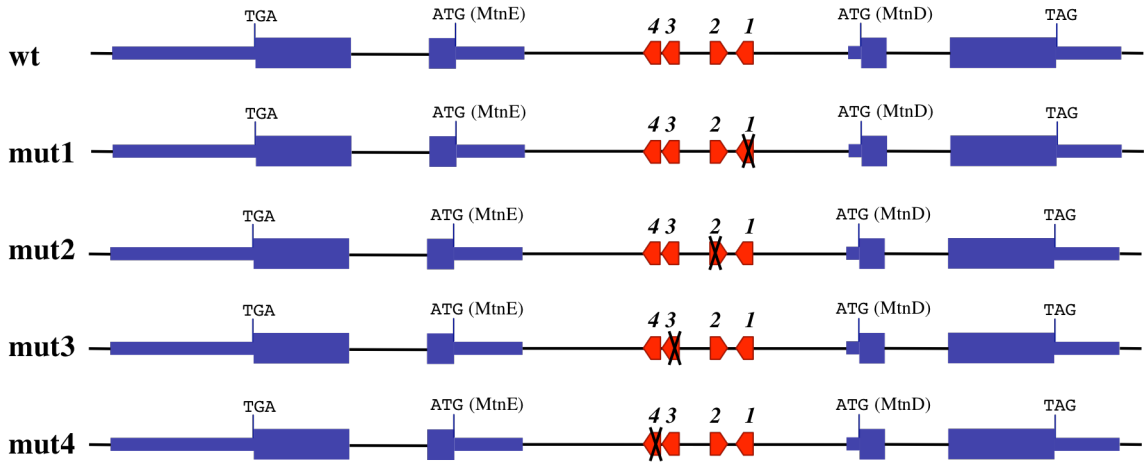
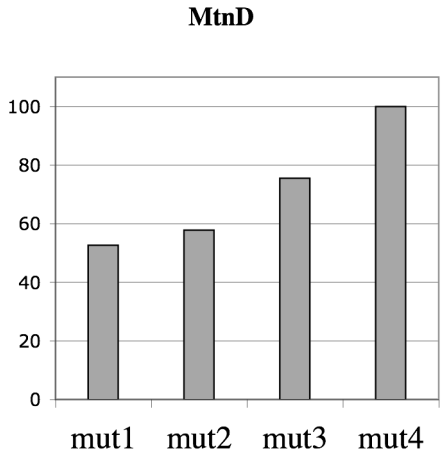


Figure 7.

A.


wt
mut1
mut2
mut3
mut4
TGA
ATG (MtnE)
4 3
2
1
ATG (MtnD)
TAG

B.


MtnD
100
80
60
40
20
0
mut1
mut2
mut3
mut4

References

1. **Andrews, G.** 2001. Cellular zinc sensors: MTF-1 regulation of gene expression. Biometals **14:**223-237.

2. **Auf der Maur, A., T. Belser, G. Elgar, O. Georgiev, and W. Schaffner.** 1999. Characterization of the transcription factor MTF-1 from the Japanese pufferfish (Fugu rubripes) reveals evolutionary conservation of heavy metal stress response. Biol Chem **380:**175-185.

3. **Bischof, J., R. Maeda, M. Hediger, F. Karch, and K. Basler.** 2007. An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. Proc Natl Acad Sci U S A **104:**3312-3317.

4. **Blindauer, C., and O. Leszczyszyn.** 2010. Metallothioneins: unparalleled diversity in structures and functions for metal ion homeostasis and more. Nat Prod Rep **27:**720-741.

5. **Brenner, R., and N. Atkinson.** 1997. Calcium-activated potassium channel gene expression in the midgut of Drosophila. Comp Biochem Physiol B Biochem Mol Biol **118:**411-420.

6. **Carter, A., B. Felber, M. Walling, M. Jubier, C. Schmidt, and D. Hamer.** 1984. Duplicated heavy metal control sequences of the mouse metallothionein-I gene. Proc Natl Acad Sci U S A **81:**7392-7396.

7. **Chiaverini, N., and M. De Ley.** 2010. Protective effect of metallothionein on oxidative stress-induced DNA damage. Free Radic Res **44:**605-613.

8. **Culotta, V., and D. Hamer.** 1989. Fine mapping of a mouse metallothionein gene metal response element. Mol Cell Biol **9:**1376-1380.

9. **Dalton, T., W. Solis, D. Nebert, and M. r. Carvan.** 2000. Characterization of the MTF-1 transcription factor from zebrafish and trout cells. Comp Biochem Physiol B Biochem Mol Biol **126:**325-335.

10. **Egli, D., A. Selvaraj, H. Yepiskoposyan, B. Zhang, E. Hafen, O. Georgiev, and W. Schaffner.** 2003. Knockout of 'metal-responsive transcription factor' MTF-1 in Drosophila by homologous recombination reveals its central role in heavy metal homeostasis. EMBO J **22:**100-108.

11. **Egli, D., H. Yepiskoposyan, A. Selvaraj, K. Balamurugan, R. Rajaram, A. Simons, G. Multhaup, S. Mettler, A. Vardanyan, O. Georgiev, and W. Schaffner.** 2006. A family knockout of all four Drosophila metallothioneins reveals a central role in copper homeostasis and detoxification. Mol Cell Biol **26:**2286-2296.

12. **Filshie, B., D. Poulson, and D. Waterhouse.** 1971. Ultrastructure of the copper-accumulating region of the Drosophila larval midgut. Tissue Cell **3:**77-102.

13. **Fish, M., A. Groth, M. Calos, and R. Nusse.** 2007. Creating transgenic Drosophila by microinjecting the site-specific phiC31 integrase mRNA and a transgene-containing donor plasmid. Nat Protoc **2:**2325-2331.

14. **Giedroc, D., X. Chen, and J. Apuy.** 2001. Metal response element (MRE)-binding transcription factor-1 (MTF-1): structure, function, and regulation. Antioxid Redox Signal **3:**577-596.

15. **Günes, C., R. Heuchel, O. Georgiev, K. Müller, P. Lichtlen, H. Blüthmann, S. Marino, A. Aguzzi, and W. Schaffner.** 1998. Embryonic lethality and liver degeneration in mice lacking the metal-responsive transcriptional activator MTF-1. EMBO J **17:**2846-2854.

16. **Hamer, D., D. Thiele, and J. Lemontt.** 1985. Function and autoregulation of yeast copperthionein. Science **228:**685-690.

17. **Heuchel, R., F. Radtke, O. Georgiev, G. Stark, M. Aguet, and W. Schaffner.** 1994. The transcription factor MTF-1 is essential for basal and heavy metal-induced metallothionein gene expression. EMBO J **13:**2870-2875.

18. **Lichtlen, P., and W. Schaffner.** 2001. Putting its fingers on stressful situations: the heavy metal-regulatory transcription factor MTF-1. Bioessays **23:**1010-1017.

19. **Margoshes M., and V. B. L.** 1957. A cadmium protein from equine kidney cortex, p. 4813-4814, J. Am. Chem. Soc.

20. **Murphy, B., G. Andrews, D. Bittel, D. Discher, J. McCue, C. Green, M. Yanovsky, A. Giaccia, R. Sutherland, K. Laderoute, and K. Webster.** 1999. Activation of metallothionein gene expression by hypoxia involves metal response elements and metal transcription factor-1. Cancer Res **59:**1315-1322.

21. **Palmiter, R.** 1998. The elusive function of metallothioneins. Proc Natl Acad Sci U S A **95:**8428-8430.

22. **Radtke, F., R. Heuchel, O. Georgiev, M. Hergersberg, M. Gariglio, Z. Dembic, and W. Schaffner.** 1993. Cloned transcription factor MTF-1 activates the mouse metallothionein I promoter. EMBO J **12:**1355-1362.

23. **Robinson, N., S. Whitehall, and J. Cavet.** 2001. Microbial metallothioneins. Adv Microb Physiol **44:**183-213.

24. **Sato, M., and I. Bremner.** 1993. Oxygen free radicals and metallothionein. Free Radic Biol Med **14:**325-337.

25. **Sato, M., T. Kawakami, M. Kondoh, M. Takiguchi, Y. Kadota, S. Himeno, and S. Suzuki.** 2010. Development of high-fat-diet-induced obesity in female metallothionein-null mice. FASEB J **24:**2375-2384.

26. **Searle, P., G. Stuart, and R. Palmiter.** 1985. Building a metal-responsive promoter with synthetic regulatory elements. Mol Cell Biol **5:**1480-1489.

27. **Searle, P., G. Stuart, and R. Palmiter.** 1987. Metal regulatory elements of the mouse metallothionein-I gene. Experientia Suppl **52:**407-414.

28. **Sigel A., Sigel H, and Sigel H.** 2009. Metal Ions in Life Sciences. RSC publishing.

29. **Simpkins, C.** 2000. Metallothionein in human disease. Cell Mol Biol **46:**465-488.

30. **Stillman, M. J.** 1995. Metallothioneins, p. 461-511, **Coordination Chemistry Reviews**, vol. 144.

31. **Stuart, G., P. Searle, H. Chen, R. Brinster, and R. Palmiter.** 1984. A 12-base-pair DNA motif that is repeated several times in metallothionein gene promoters confers metal regulation to a heterologous gene. Proc Natl Acad Sci U S A **81:**7318-7322.

32. **Stuart, G., P. Searle, and R. Palmiter.** 1985. Identification of multiple metal regulatory elements in mouse metallothionein-I promoter by assaying synthetic sequences. Nature **317:**828-831.

33. **Wang, Y., I. Lorenzi, O. Georgiev, and W. Schaffner.** 2004. Metal-responsive transcription factor-1 (MTF-1) selects different types of metal response elements at low vs. high zinc concentration. Biol Chem **385:**623-632.

34. **Weaver, R., and C. Weissmann.** 1979. Mapping of RNA by a modification of the Berk-Sharp procedure: the 5' termini of 15 S beta-globin mRNA precursor and mature 10 s beta-globin mRNA have identical map coordinates. Nucleic Acids Res **7:**1175-1193.

35. **Westin, G., T. Gerster, M. Müller, G. Schaffner, and W. Schaffner.** 1987. OVEC, a versatile system to study transcription in mammalian cells and cell-free extracts. Nucleic Acids Res **15:**6787-6798.

36. **Westin, G., and W. Schaffner.** 1988. A zinc-responsive factor interacts with a metal-regulated enhancer element (MRE) of the mouse metallothionein-I gene. EMBO J **7:**3763-3770.

37. **Winge, D., K. Nielson, W. Gray, and D. Hamer.** 1985. Yeast metallothionein. Sequence and metal-binding properties. J Biol Chem **260:**14464-14470.

BMC Developmental Biology

Research article

Dumpy-30 family members as determinants of male fertility and interaction partners of metal-responsive transcription factor 1 (MTF-1) in *Drosophila*

Alla Vardanyan[1], Lilit Atanesyan[1], Dieter Egli[1], Sunil Jayaramaiah Raja[2], Monica Steinmann-Zwicky[3], Renate Renkawitz-Pohl[2], Oleg Georgiev[1] and Walter Schaffner*[1]

Address: [1]Institute of Molecular Biology, University of Zurich, Winterthurer St. 190, CH-8057 Zurich, Switzerland, [2]Philipps-University, Dept. Developmental Biology, Karl-von-Frisch-Str. 8, D-35043 Marburg, Germany and [3]Zoological Institute, University of Zurich, Winterthurer St. 190, CH-8057 Zurich, Switzerland

Email: Alla Vardanyan - alla.vardanyan@molbio.unizh.ch; Lilit Atanesyan - lilit.atanesyan@molbio.uzh.ch; Dieter Egli - degli@mcb.harvard.edu; Sunil Jayaramaiah Raja - raja@embl.de; Monica Steinmann-Zwicky - mstzw@zool.uzh.ch; Renate Renkawitz-Pohl - renkawit@staff.uni-marburg.de; Oleg Georgiev - Ole@molbio.uzh.ch; Walter Schaffner* - walter.schaffner@molbio.uzh.ch

* Corresponding author

Published: 27 June 2008

BMC Developmental Biology 2008, **8**:68 doi:10.1186/1471-213X-8-68

Received: 10 April 2008
Accepted: 27 June 2008

This article is available from: http://www.biomedcentral.com/1471-213X/8/68

Abstract

Background: Metal-responsive transcription factor 1 (MTF-1), which binds to metal response elements (MREs), plays a central role in transition metal detoxification and homeostasis. A *Drosophila* interactome analysis revealed two candidate dMTF-1 interactors, both of which are related to the small regulatory protein Dumpy-30 (Dpy-30) of the worm *C. elegans*. Dpy-30 is the founding member of a protein family involved in chromatin modifications, notably histone methylation. Mutants affect mating type in yeast and male mating in *C. elegans*.

Results: Constitutive expression of the stronger interactor, Dpy-30L1 (CG6444), in transgenic flies inhibits MTF-1 activity and results in elevated sensitivity to Cd(II) and Zn(II), an effect that could be rescued by co-overexpression of dMTF-1. Electrophoretic mobility shift assays (EMSA) suggest that Dpy-30L1 interferes with the binding of MTF-1 to its cognate MRE binding site. Dpy-30L1 is expressed in the larval brain, gonads, imaginal discs, salivary glands and in the brain, testes, ovaries and salivary glands of adult flies. Expression of the second interactor, Dpy-30L2 (CG11591), is restricted to larval male gonads, and to the testes of adult males. Consistent with these findings, *dpy-30*-like transcripts are also prominently expressed in mouse testes. Targeted gene disruption by homologous recombination revealed that *dpy-30L1* knockout flies are viable and show no overt disruption of metal homeostasis. In contrast, the knockout of the male-specific *dpy-30L2* gene results in male sterility, as does the double knockout of *dpy-30L1* and *dpy-30L2*. A closer inspection showed that Dpy-30L2 is expressed in elongated spermatids but not in early or mature sperm. Mutant sperm had impaired motility and failed to accumulate in sperm storage organs of females.

Conclusion: Our studies help to elucidate the physiological roles of the Dumpy-30 proteins, which are conserved from yeast to humans and typically act in concert with other nuclear proteins to modify chromatin structure and gene expression. The results from these studies reveal an inhibitory effect of Dpy-30L1 on MTF-1 and an essential role for Dpy-30L2 in male fertility.

Background

Metal-responsive transcription factor 1 (MTF-1) can cooperate, in a positive or negative manner, with other transcription factors binding to their own DNA sites nearby (USF1, [1]; NFI, [2,3]; Sp1, [4]; NF-kB [5]), but no MTF-1-specific coactivators or corepressors were described so far. A general interaction analysis of *Drosophila* proteins by means of the yeast two-hybrid system [6] revealed two closely related proteins as potential interaction partners of MTF-1 (see below). These interaction proteins were encoded by genes designated CG6444 and CG11591 [7]. Both belong to a protein family that is conserved from yeast to humans and whose founding member was described in the nematode *C. elegans* as Dumpy-30 (Dpy-30), a protein involved in dosage compensation of sex chromosomes [8]. Dpy-30 is required for sex-specific association of Dpy-27, a chromosome condensation protein homolog, with the hermaphrodite's X chromosomes. Besides causing XX-specific lethality, the *dpy-30* mutation in XO animals causes developmental delay, small body size, inability to mate and abnormal tail morphology [9]. These phenotypes suggest an involvement of Dpy-30 also in processes other than dosage compensation. The yeast homolog of *C. elegans* Dpy-30, Sdc1, was identified as an important component of the eight-member complex (SET1C protein complex), which functions as a histone 3 lysine 4 (H3-K4) methyltransferase [10]. The loss of individual SET1 protein complex subunits differentially affects SET1 stability, complex integrity and the distribution of H3K4 methylation along active genes. Such mutations cause defects in maintenance of telomere length [11] and in DNA repair [12,13]. Dpy-30 and its close relatives contain a short motif related to the dimerization motif in the regulatory subunit of Protein Kinase A. This motif consists of two α-helices that form a special type of four-helix bundle during dimerization [14]. Until recently no data were available on one of the *Drosophila* homologs, CG6444, while the other, CG11591, was shown to be expressed in testes by genome-wide microarray analysis of transcription [15].

As mentioned, the interaction partner of Dpy-30-like proteins in the *Drosophila* interaction study was identified as metal-responsive transcription factor 1 (MTF-1). MTF-1 is a key regulator of heavy metal homeostasis and detoxification in higher eukaryotes [16-19]. In mammals, MTF-1 controls a number of genes for metal homeostasis and is also essential for embryonic liver development [20-23].

MTF-1 binds via its zinc fingers to metal-responsive elements (MREs) in the promoter/enhancer region of target genes [16,24] and activates their transcription. Metallothioneins are the best characterized target genes of MTF-1; they encode small, cysteine-rich proteins with an ability to scavenge excess heavy metal ions [25-27]. *Drosophilae* mutant for dMTF-1, the homolog of mammalian MTF-1, are viable but more sensitive to elevated concentrations of heavy metals, as well as to copper scarcity [28]. Upon copper starvation, dMTF-1 activates transcription of the gene encoding Ctr1B, a high affinity copper importer [29]. Recently several additional target genes of MTF-1 in mammals and in *Drosophila* were identified and characterized in our laboratory by microarray and specific transcript analysis [30,31] but little is known to date about proteins interacting with and/or regulating *Drosophila* MTF-1 function.

Here we show that transgenes of both *Drosophila* Dpy-30 orthologs, CG6444 and CG11591, hereafter termed Dpy-30-like 1 (Dpy-30L1) and Dpy-30-like 2 (Dpy-30L2), respectively, inhibit MTF-1-dependent reporter gene expression in cell culture. Constitutive expression of a Dpy-30L1 transgene in flies results in elevated sensitivity to Cd(II) and Zn(II), while Dpy-30L2 overexpression has no such effect. Consistent with metal resistance, only the Dpy-30L1 transgene inhibited dMTF-1 activity in flies. Gene knockout by homologous recombination revealed that *dpy-30L1* null mutant flies are viable and fertile and maintain a seemingly normal metal homeostasis, while knockout of the male-specific *dpy-30L2* results in male sterility. Sperm motility in *dpy-30L2* mutants is impaired and drastically decreases with age. After mating mutant sperm is transferred to the uterus but does not accumulate in the seminal receptacle and spermathecae, making successful fertilization impossible. These findings reveal a major role of Dpy-30 proteins in male fertility and sperm motility.

Results

Inhibition of MTF-1-dependent reporter expression in Drosophila *Schneider S2 cells*

The *Drosophila* interactome study of Giot *et al.* [6] had revealed three proteins that display very good (Dpy-30L1), good (Dpy-30L2), and weak (CG11061) interaction with dMTF-1. In order to characterize the role of these proteins in *Drosophila*, especially in the context of metal homeostasis, the open reading frames (Figure 1) of all three were cloned into a *Drosophila* expression vector and analyzed by transfection and co-transfection studies in insect cells. The third protein reported to interact with dMTF-1 only weakly, CG11061, was listed as a protein putatively involved in Golgi organization and biogenesis, mitosis and protein targeting. In our hands it did not affect MTF-1 function (data not shown), thus rendering doubtful a physiological relevance of the predicted interaction.

In *Drosophila* Schneider S2 cells [32], transfection of *dpy-30L1* or *dpy-30L2* inhibited the expression of MTF-1-dependent reporter genes driven either by the promoter of the *Drosophila* metallothionein A (MtnA) (not shown) or

Figure 1

Alignment of *Drosophila* Dpy-30L1 and Dpy-30L2 with their orthologs from different species. The core domain of these short proteins is highly conserved among different species. Grey shaded: similar, black shaded: identical aa. Dpy-30 " core" indicates the consensus core sequence.

by a synthetic promoter consisting of four tandem metal response elements (MREs), the binding sites of MTF-1 (Figure 2A). The effect on the synthetic MRE promoter was more pronounced, suggesting that Dpy-30L1 and Dpy-30L2 indeed interact with dMTF-1 and thereby interfere with its activity, as MTF-1 is the only factor known to bind MREs. This inhibitory effect could also be shown in the whole organism expressing an YFP reporter gene driven by the metallothionein (MtnA) promoter (Figure 2B). Here, the response (YFP expression) to copper, and especially to cadmium, was strongly reduced, whereas under copper starvation conditions no difference was observed.

Due to the high degree of conservation among all the members of this protein family, we screened the mouse and human genome for the orthologs (Figure 1) and subcloned the Dpy-30-like members both from mouse and human. These mammalian Dpy-30-like proteins were as effective as the *Drosophila* proteins in repressing dMTF-1 activity in *Drosophila* Schneider cells (Figure 3B and not shown). However, it appears that the antagonistic interaction between MTF-1 and Dpy-30 family members is specific to *Drosophila* MTF-1: mammalian MTF-1 was not affected by Dpy-30-type proteins, irrespective of whether the test was done in *Drosophila* cells (Figure 3) or mammalian cells (not shown).

To gain further insights to the inhibitory effect of Dpy-30L1 on dMTF-1, we did an electrophoretic mobility shift assay (EMSA) of transfected VSV-tagged MTF-1, without or with co-transfected *dpy-30L1*. The reduced band intensity of the shifted MRE oligo suggests that Dpy-30L1 interferes with binding of MTF-1 to its cognate MRE DNA (Figure 4).

Flies overexpressing Dpy-30L1 are sensitive to heavy metal load

We generated transgenic flies with ubiquitous, constitutive expression of *dpy-30L1* or *dpy-30L2*, taking advantage of the UAS-Gal4 system whereby Gal4 was driven by the *Drosophila* actin promoter. Flies overexpressing Dpy-30L1 were raised during their entire development on normal food, or food supplemented with different heavy metals. They did not show a phenotype when kept on standard food but were much more sensitive to heavy metal load, especially to cadmium and zinc, while sensitivity to copper was only marginally affected (Figure 5). The sensitivity to cadmium and zinc could be rescued by co-overexpression of an MTF-1 transgene (Figure 5). This shows that the

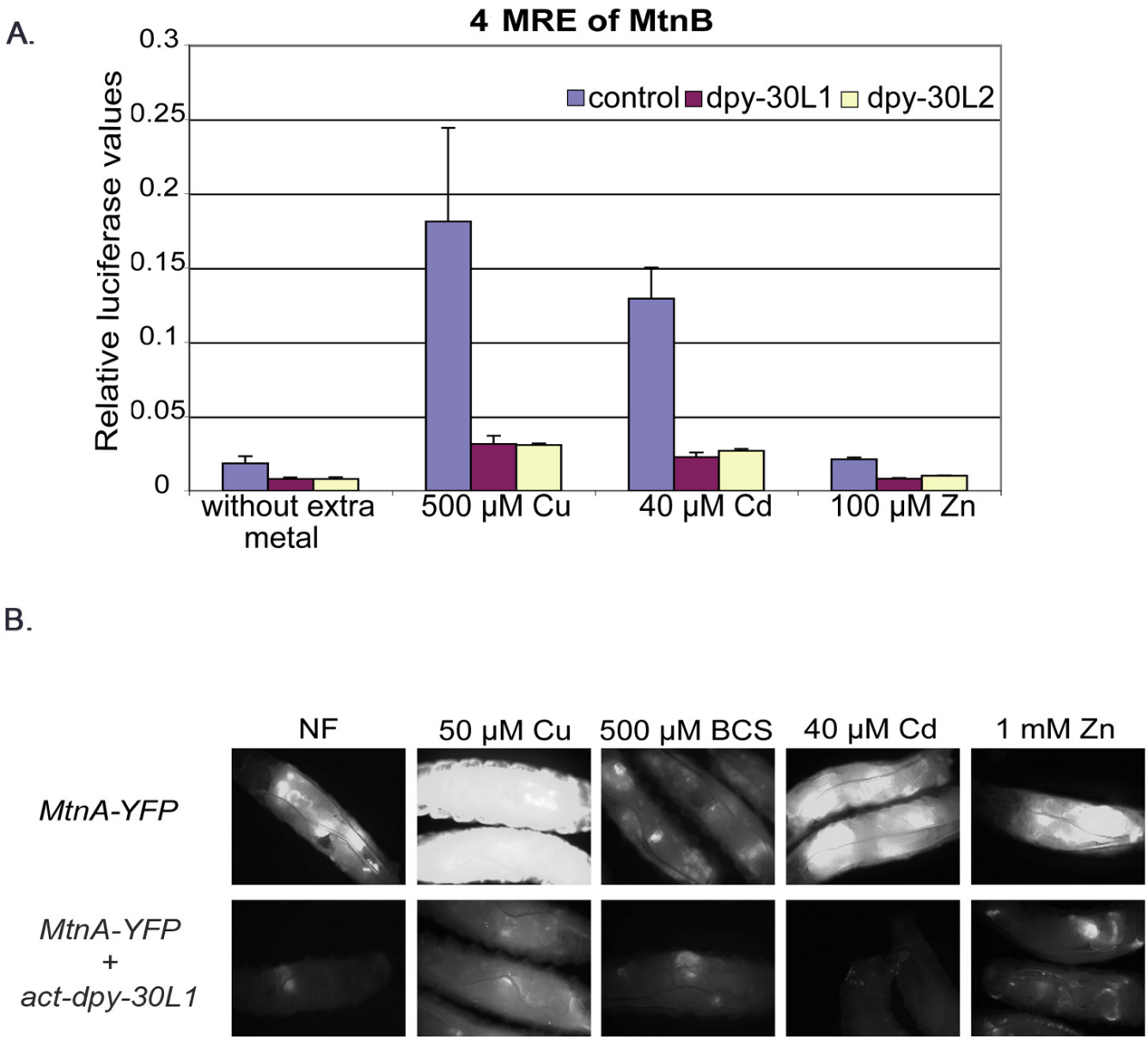


Figure 2

Effect of Dpy-30L1 overexpression on MTF-1 dependent reporters in cell culture and *in vivo*. A) In transiently transfected *Drosophila* Schneider S2 cells [32], the ratio of firefly (reporter) to renilla (reference) luciferase activity is shown. Reporter: 4xMRE from the metallothionein B (MtnB) promoter [18] fused to firefly luciferase; reference: tubulin promoter fused to renilla luciferase [53]. Dpy-30L1 and Dpy-30L2 expression constructs were under the control of the actin promoter. 72 hours after transfection, cells were treated with the indicated concentrations of heavy metals for 24 hours. B) Expression level of green fluorescent protein in transgenic larvae that carry an MtnA-YFP reporter construct. Transgenic flies were allowed to lay eggs on normal food or food supplemented with different heavy metals.

metal sensitivity of organisms expressing transgenic Dpy-30L1 was not merely reflecting a generally lower resistance to stress but rather a disturbed metal-specific stress response. This point was corroborated by raising flies in excess iron, a metal that is handled by a pathway different from the MTF-1/metallothionein system. Neither an increased sensitivity nor a rescue effect could be observed upon overexpression of Dpy-30L1 and/or dMTF-1 (data not shown). Although both Dpy-30L1 and Dpy-30L2 overexpression inhibited MTF-1 function in cell culture, only Dpy-30L1 was effective in a transgenic fly. This leads to the conclusion that there are functional differences

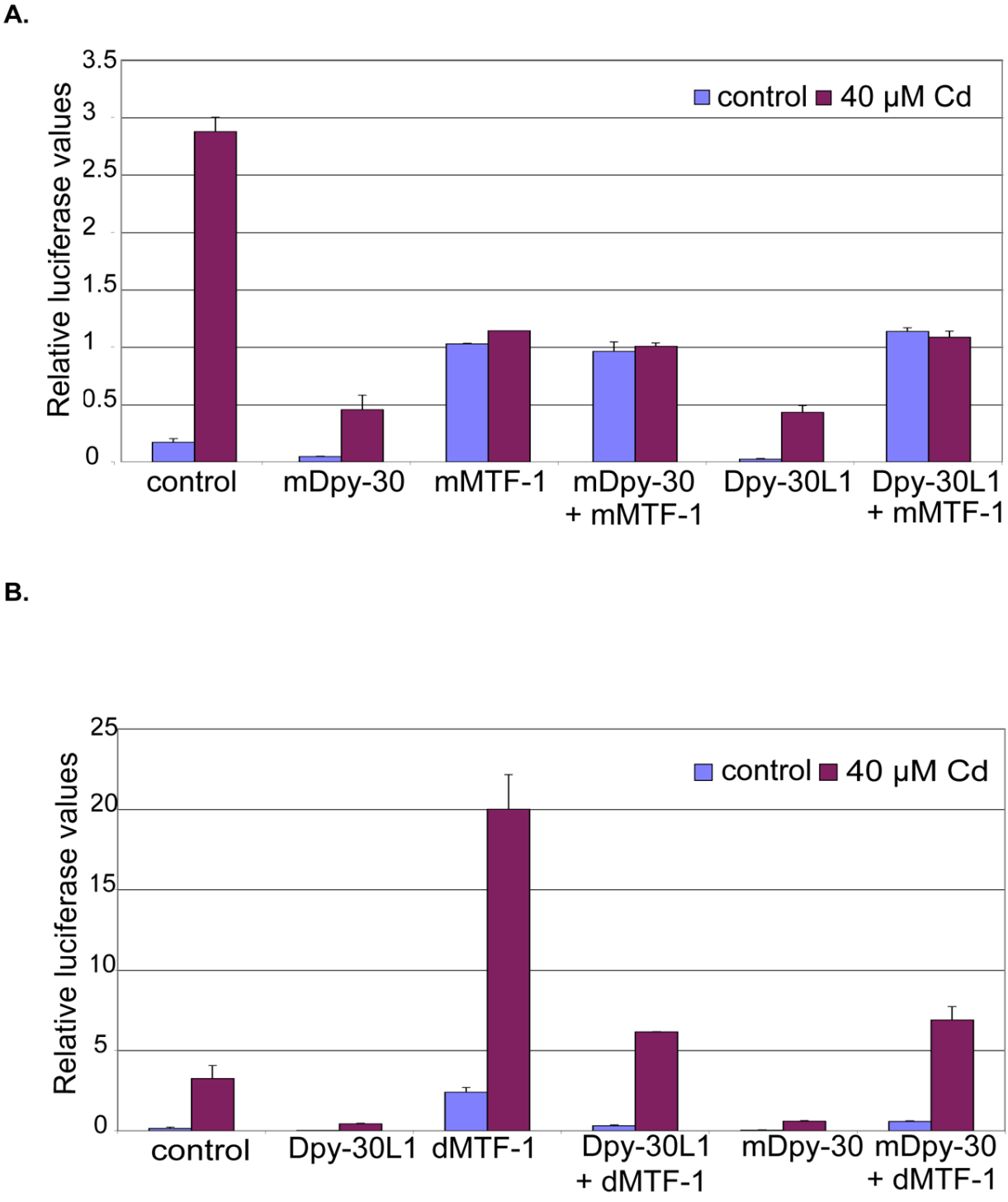


Figure 3

Inhibitory effect of Dpy-30L1, Dpy-30L2 and their mammalian orthologs is restricted to *Drosophila* MTF-1. The ratio of firefly to renilla luciferase activity in transiently transfected *Drosophila* Schneider S2 cells is shown. Reporter, MtnA promoter fused to firefly luciferase; reference, tubulin promoter fused to renilla luciferase. Dpy-30L1 and the mouse ortholog were under the control of the actin promoter. 72 hours after transfection, the medium in half of the plates was supplemented with 40 μM cadmium chloride for 24 hours, while the others served as controls. A) Mouse MTF-1 was co-transfected in the indicated samples; B) *Drosophila* MTF-1 was co-transfected in the indicated samples.

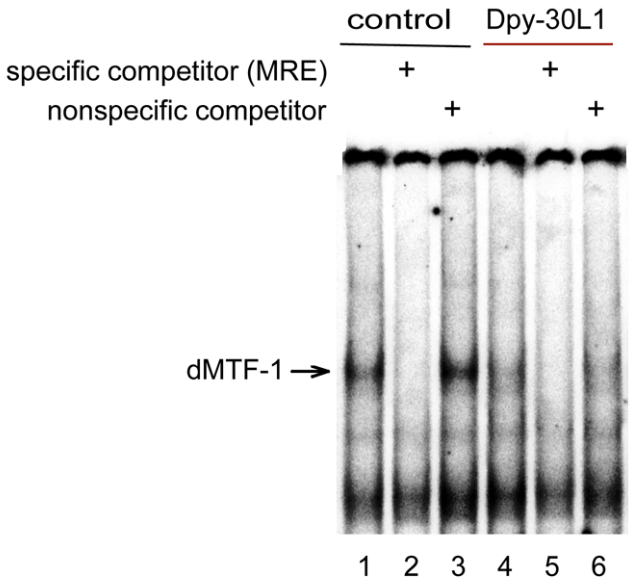


Figure 4

Reduced DNA binding of dMTF-1 upon co-expression of Dpy-30L1. DNA binding by dMTF-1 was determined by EMSA. *Drosophila* S2 cells were transfected with VSV tagged dMTF-1 expression plasmids, and 20 μg of nuclear protein extract was used for each bandshift reaction. Lanes 1 and 4, bandshift with [^{32}P]-labeled MRE consensus oligonucleotide (MRE-s). Lanes 2 and 5, same conditions but also including a 200-fold excess of unlabeled MRE-s competitor oligonucleotide (specific competitor). Lanes 3 and 6, same conditions but with a 200-fold excess of unlabeled Sp1 oligonucleotide (nonspecific competitor). Cells had been treated for 6 hours with medium containing 100 μM zinc sulfate.

between the two related proteins that become evident only in whole-organism studies.

Expression pattern of Dpy-30L1 and Dpy-30L2

To determine the expression pattern of both genes in larvae and adult flies, transgenes were constructed where a fluorescent reporter (YFP) was under the control of approximately 7 kb of genomic region from Dpy-30L1 or Dpy-30L2. The expression pattern of the two genes was quite distinct: The *dpy-30L1* regulatory region induced expression in multiple larval tissues, notably brain, gonads, imaginal discs and salivary glands. In adult flies, expression was seen in the brain, testes, ovaries and salivary glands. In contrast, the expression of Dpy-30L2-YFP was confined exclusively to male gonads in larvae, and to the testes in adult flies. Further dissection of the expression pattern of Dpy-30L2-YFP during spermatogenesis revealed that Dpy-30L2-YFP is expressed in elongated spermatids at the "canoe-like" stage but not during the early stages of spermatogenesis or in mature sperm (not shown), which is also consistent with the online *Drosophila* testis gene expression database [33]. The expression pattern of these transgenic constructs was very similar to the ones derived from a genome-wide transcription map recently published online in Flyatlas [34,35].

Targeted gene disruption shows that Dpy-30L2 is essential for male fertility

In order to determine the *in vivo* role of the two Dpy-30-like proteins, we disrupted both of the corresponding genes by means of homologous recombination [36]. Somewhat unexpectedly, *dpy-30L1* knockout flies turned out to be viable and fertile under laboratory conditions and did not show any obvious alterations in metal resistance/sensitivity phenotypes (data not shown). In the mutated locus, the yellow fluorescent protein (YFP) and SV40 polyadenylation/termination sequence was followed, out-of-frame, by a truncated Dpy-30L1 coding sequence (for details, see Materials & Methods). Though unlikely, we cannot rule out the possibility that the residual Dpy-30L1 sequence was expressed by re-initiation of transcription and translation from an internal site in the coding region, thus producing a hypomorph, rather than a null mutation. In contrast, disruption of the male-specific *dpy-30L2* gene resulted in complete male sterility. Combination of the two mutations did not reveal any additional phenotypic features, i.e., male flies were again sterile but otherwise normal under laboratory conditions.

We attempted to identify more precisely the defects in the reproductive system of *dpy-30L2* knockout males. A dissection of *dpy-30L2*$^{6\text{-}1}$ males revealed apparently normal testes that contained sperm. In the nematode *C. elegans*, Dpy-30 is known to be involved in dosage compensation, a process that equalizes the expression of X-chromosomes in XX and XO animals [8], and in yeast it methylates histones [38,39]. Dpy-30L2 is specifically expressed in the spermatid stage where transcription is repressed in germ cells and histones are removed from DNA to be replaced by protamines. Thus we wondered whether in *Drosophila*, loss of Dpy-30L2 distorts chromatin structure at this critical stage. However, loss of histone H2A variant D (H2AvD) expression, a hallmark of the transition to the protamine-loaded sperm, was not affected (Figure 6), and also the protamine B-eGFP distribution pattern was not disturbed. Furthermore, the marker Mst77F-eGFP was inconspicuous in that it was associated with DNA at the appropriate stage of spermatogenesis (Figure 6). Mst77F is a distant relative of the linker histone H1/H5 family and has been proposed to support the transition to compact *Drosophila* sperm chromatin [39,40]. Unlike its mammalian homolog (mHILS1), Drosophila Mst77F persists in mature sperm nuclei [40].

However, we found one clear difference between mutant and wild type flies: dissection of the reproductive tract of females that had mated with dpy-30L2 mutant males

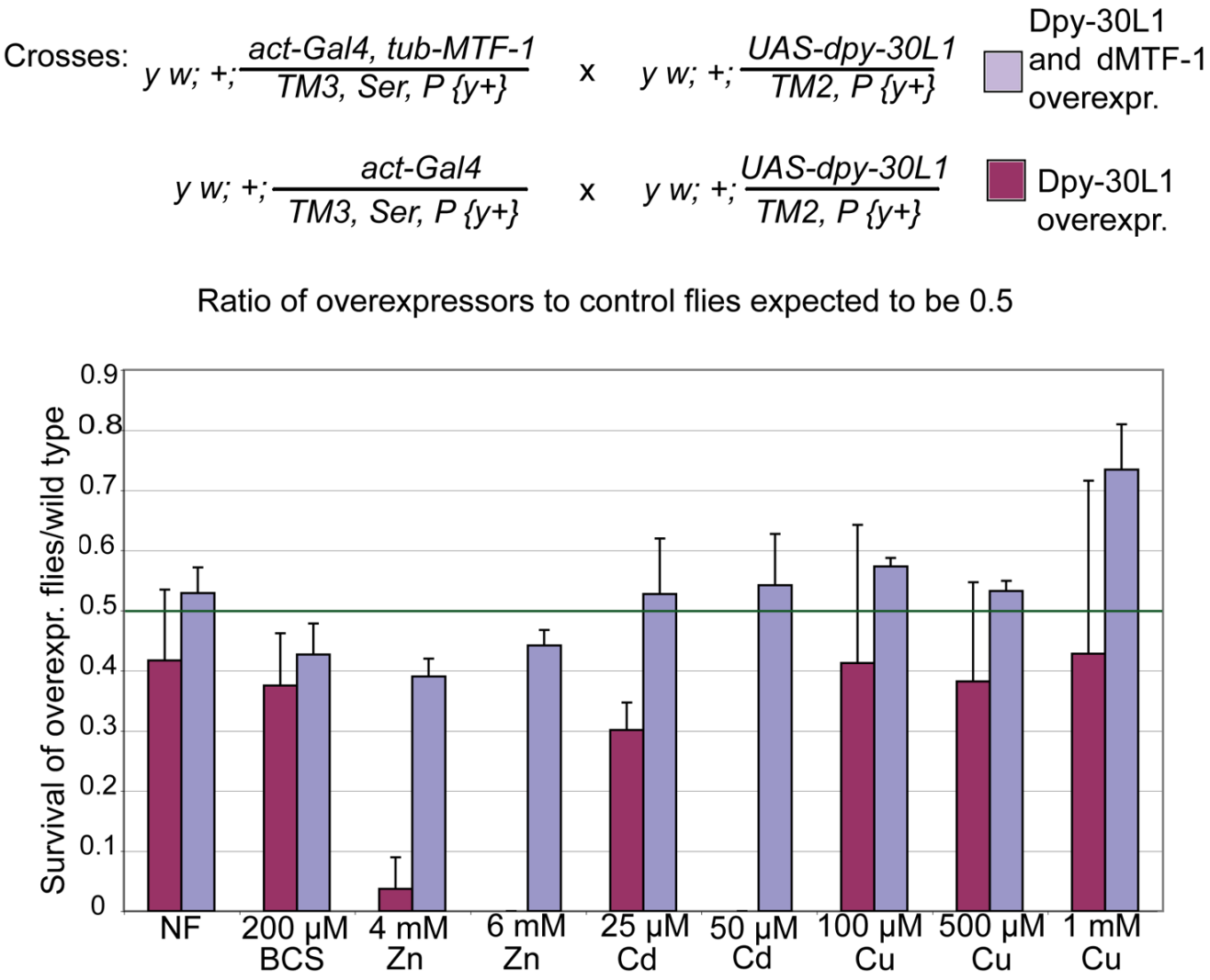


Figure 5

Sensitivity of *Drosophila* to heavy metal load. Crosses of flies with the indicated genotypes were done on normal food (NF) or on food supplemented with the indicated metals. Flies were allowed to lay approximately the same amount of eggs, then in each tube the ratio of eclosed Dpy-30L1 overexpressors to controls was determined. In normal food, the cross was expected to yield 1/3 overexpressors and 2/3 controls i.e., a ratio of 0.5, which is indicated by a green line. act – actin, tub – tubulin.

revealed that sperm were confined to the uterus, which means that they had failed to be transmitted to the seminal receptacle and the spermathecae (Figure 7). Since from these latter sites sperm are used to fertilize eggs, this mislocalization could, in part or completely, explain the sterility of mutant males. A possible reason for mislocalization of dpy-30L2 mutant sperm in females could be impaired or uncoordinated motility of sperm. Dissection of the reproductive tract of females that had been mated either with Oregon R or with *dpy-30L2*$^{6\text{-}1}$ males showed that *dpy-30L2* knockout sperm indeed lose their motility after transfer to the female reproductive tract (Table 1). More detailed analysis of sperm amount and motility in males revealed an age-dependent decrease in both amount and motility of *dpy-30L2*$^{6\text{-}1}$ sperm, with complete loss of motility in 20-day-old males in contrast to heterozygous males. Taken together, these results reveal that Dpy-30L2 is important for sperm production and motility.

Discussion

In transfected cells, both of the Dpy-30 orthologs of *Drosophila*, termed Dpy-30L1 and Dpy-30L2 (Dumpy-30-like1 and -like2), inhibit the activity of MTF-1 (metal-

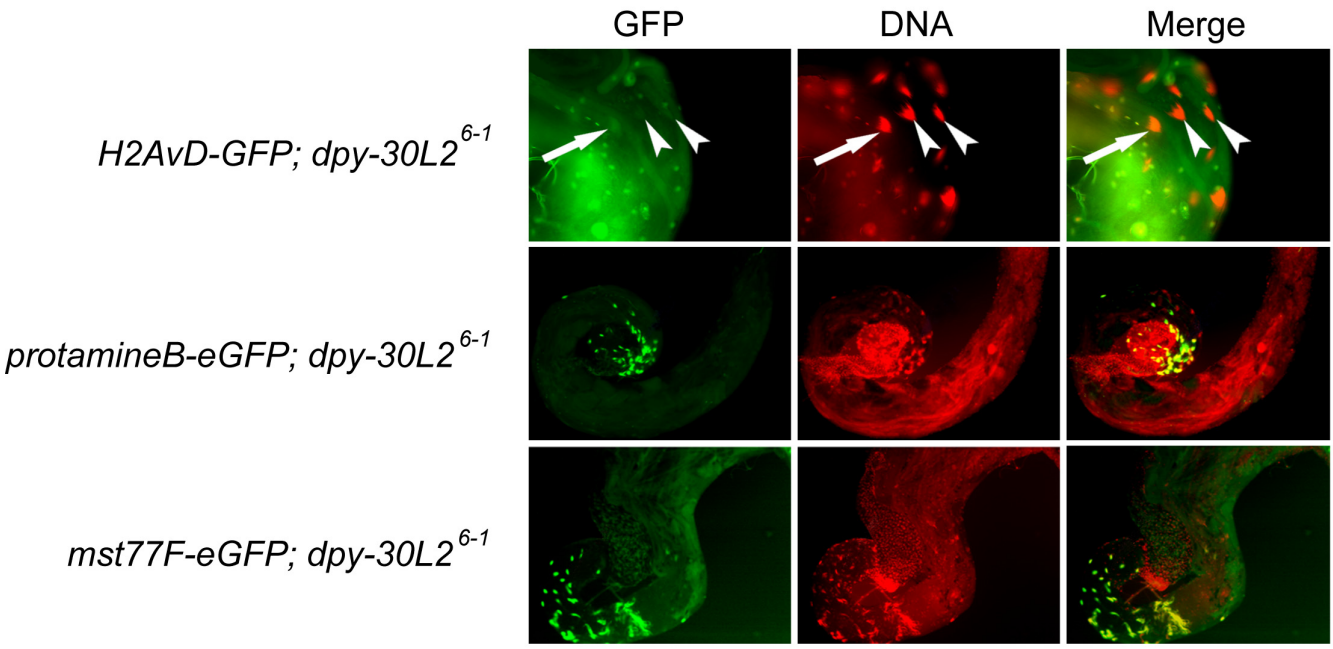


Figure 6
Transition of histones to protamines in *dpy-30L2* knockout males. Degradation of histones was checked in *dpy-30L2* knockout (*dpy-30L2*$^{6\text{-}1}$) males carrying a transgene with a fusion of GFP to the coding sequence of the histone H2A D variant (H2AvD). Arrow, H2AvD-GFP in degradation; arrowheads, H2AvD-GFP degraded. Incorporation of protamine B and Mst77F was analyzed in *dpy-30L2* knockout males that carry either a transgene of protamine B fused to eGFP or of Mst77F fused of eGFP. During the "canoe" and "post-canoe" stages of spermatid development, ProtamineB-eGFP and Mst77F-eGFP incorporation in the spermatid nucleus appeared to be normal in *dpy-30L2* knockout males. Any (diffuse) YFP signal from the *dpy-30L2* promoter was not filtered out.

responsive transcription factor 1), while in transgenic flies, such an effect was only seen with the stronger interactor Dpy-30L1. Consistent with such an inhibition, transgenic flies were sensitive to cadmium or zinc load, while copper sensitivity was only marginally affected. The increased metal sensitivity could be rescued by co-overexpression of dMTF-1. An EMSA assay revealed a weakened binding of MTF-1 to MRE DNA in the presence of Dpy-30L1. Taken together, these results suggest that for detoxification of Cd(II) or Zn(II) a higher level of MTF-1 is required than for Cu(II) detoxification. Studies with partial inactivation mutants of dMTF-1 are in agreement with such a notion (A.V., H. Yepiskoposyan and W.S., unpublished). Unexpectedly, only MTF-1 of insect origin responded to Dpy-30 type proteins: while the human and mouse Dpy-30 members also inhibited *Drosophila* MTF-1 across species boundaries, activity of human MTF-1 was unchanged in the presence of either *Drosophila* or mammalian Dpy-30 members. This indicates some degree of functional divergence between *Drosophila* and mammalian MTF-1 during evolution, in spite of a conserved role in heavy metal homeostasis and detoxification. We consider the Dpy-30-dMTF-1 interactions observed in the interactome study [6] relevant because (i) the two major interactors Dpy-30L1 and L2 are members of the same protein family; (ii) a (negative) functional interaction with dMTF-1 was seen with both of them, and with their mammalian Dpy-30 homolog, in transfected cells; (iii) Dpy-30L1, the stronger interactor, also produced an effect *in vivo*, and (iv) it inhibited the binding of dMTF-1 to its cognate DNA sequence.

As a complement to transgenic expression of Dpy-30L1 and Dpy-30L2, we also tested a loss of function of the two proteins. Disruption of short genes in *Drosophila* has been a great challenge since small targets are rarely hit by random mutagenesis. To circumvent this problem, we eliminated Dpy-30L1 and L2 function separately by homologous recombination [36,41]. Somewhat unexpectedly, knockout of neither Dpy-30L1 nor Dpy-30L2 affected metal handling under the conditions tested, but Dpy-30L2 which is specifically expressed in male gonads, turned out to be essential for male fertility.

Sdc1, the yeast homolog of Dpy-30, is a component of SET1C, also called COMPASS (*com*plex *p*roteins *ass*ociated

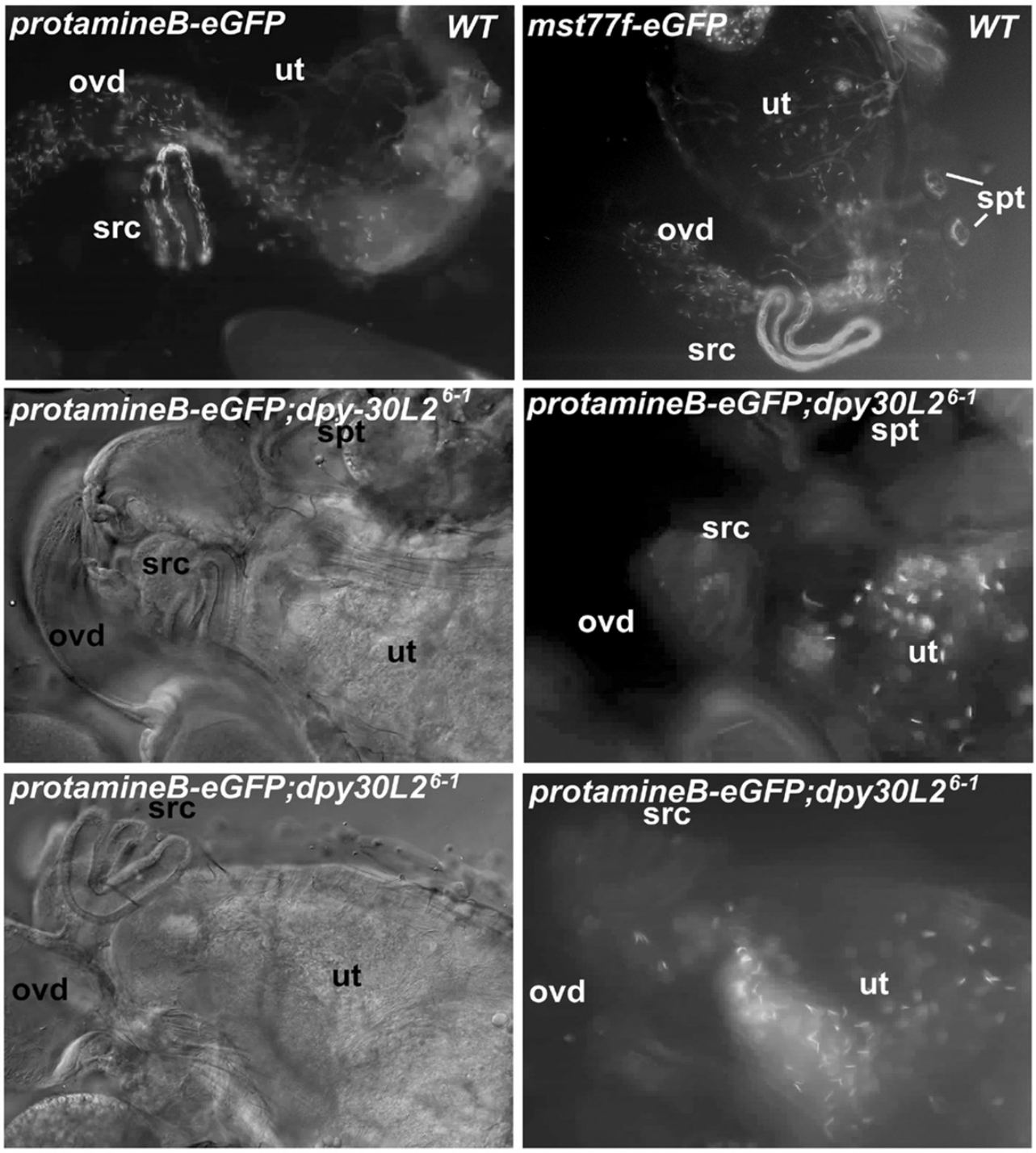


Figure 7
Dpy-30L2 knockout sperm in the wild type female reproductive system. Sperm of wild type males were marked by Mst77F or protamineB fused to eGFP to follow its fate in the female reproductive system. Wild type females were allowed to mate with wild type or mutant males, the females were then dissected and checked for a GFP signal in their reproductive system. 30 minutes after mating, 20% of wild type sperm had accumulated in female storage organs. However, mutant sperm remained in the uterus and failed to be transferred to seminal receptacle and spermathecae, the female sperm storage organs. Ov – ovaries; spt – spermatheca; src – seminal receptacle; ut – uterus; ovd – oviduct.

Table 1: Sperm presence and motility in female reproductive tract

Male genotype	Female genotype	Days after cross	Number of tested females	Number of females with sperm in seminal receptacle		Number of females with sperm in uterus/oviduct	
				Sperm	Motile	Sperm	Motile
Oregon R	Oregon R	5	6	5	5	1	No
	*XY hs-tra**	5	6	4	4	4	1
	*XY hs-tra**	1	6	6	6	4	4
dpy-30L2$^{6-1}$	Oregon R	5	6	0	-	0	-
	*XY hs-tra**	5	10	0	-	3	No
	*XY hs-tra**	1	6	0	-	6	No

*XY hs-tra** – XY fly carrying a single copy of the hs-*tra*-female plasmid [54]. These flies are females, which contain no eggs, but rather small nonfunctional germ cells in their gonads. As no eggs are laid, sperm can be detected in the uterus or in oviducts long after mating. The experiment shows that mutant sperm is not stored in the seminal receptacle and that it is immotile when remaining in the uterus or in the oviduct. In contrast, motile wildtype sperm is found both in seminal receptacles and in the uterus or oviducts of the analyzed females.

with SET1 protein). SET1C methylates histone H3 at lysine residue 4 [38]. Yeast strains mutant for SET1, although viable, display defects in cell growth, rDNA silencing [42], and silencing of telomeres and mating type loci [11]. In *C. elegans*, the dosage compensation complex (DCC), which among other proteins includes Dpy-30, represses X-chromosomal transcription in cells of XX animals. The complex binds preferentially to promoter regions and seems to be required for the early steps of dosage compensation, not for its maintenance [43]. The SET1C complex has also been shown to activate some specific genes, notably for DNA repair genes. This activation is however an indirect one, via repression of a signaling cascade [13]. Direct activation of target genes is also possible, at least in mammals: a human homolog of SET1C, the MLL (mixed-lineage leukemia) complex which also has methyltransferase activity and is ivolved in tumor cell proliferation [44], positively regulates Hox gene expression through binding to promoter sequences [45]. Recent investigations have shown that the human MLL2/ALR complex contains the human ortholog of Dpy-30 [46]. Taken together, these data indicate a conserved role of Dpy-30 family members in the modulation of chromatin structure and transcription.

However, there are clear differences as well. The *Drosophila* trithorax complex, the homolog of yeast SET1C, is essential for viability. Our findings suggest that flies lacking both Dpy-30L1 and Dpy-30L2 are viable and that Dpy-30 orthologs of *Drosophila* are not obligatory components of the trithorax complex. The only mutant phenotype we observed was male sterility in the absence of Dpy-30L2. A hallmark of spermatogenesis, the replacement of histones by protamines [47] is not affected in the Dpy-30L2 mutant. Because transcriptional silencing of the spermatid genome seems to occur independently of protamines [39], it appears still possible that Dpy-30L2 is required for proper gene silencing during spermatogenesis.

In yeast, *C. elegans* and *Drosophila*, Dpy-30 members serve different but important functions, perhaps converging, in metazoans, on sex-specific gene expression programs, compatible with the fact that the single Dpy-30 ortholog of the mouse is strongly expressed in testes.

Conclusion

Dumpy-30 (Dpy-30) type proteins are conserved from yeast to humans but their function in higher eukaryotes is only partially understood. Here we have characterized the two Dpy-30 familiy members in *Drosophila*. Strong expression of Dpy-30L1 can inhibit the activity of MTF-1 (metal-responsive transcription factor 1), resulting in elevated sensitivity of flies to cadmium and zinc load. The second member, Dpy-30L2, is only expressed in the male genital tract; targeted gene disruption of *dpy-30L2* results in male sterility associated with reduced motility of sperm and failure to be transferred to the female's seminal receptacles. Like Drosophila Dpy-30L2, the mouse Dpy-30 homolog is strongly expressed in testes, from where the expressed sequence tag (EST) was obtained [48]. Thus Dpy-30 family members may well be required for male fertility also in mammals.

Methods

Database searches and computer analysis of the sequences

Blast searches for mammalian and yeast orthologs were performed using the NCBI BLAST service. Sequence alignments were performed using ClustalW and Boxshade.

Fly food and heavy metal sensitivity assay

Flies were raised on standard cornmeal molasses-based food. For sensitivity assays, food was supplemented with $CdCl_2$, $CuSO_4$, $ZnCl_2$ or 500 μM copper chelator BCS disodium salt hydrate (Sigma-Aldrich 14, 662-5). The concentrations of trace metals in normal food, based on the content of the individual ingredients, are ~5 μM for copper and 150 μM for zinc. Flies with indicated genotypes

were allowed to lay eggs for 2 days on normal food or food supplemented with different heavy metals, and eclosed flies were counted. *Drosophila* cultures were kept at the standard temperature of 25°C.

Targeted gene disruption by homologous recombination

The targeting construct of the *dpy-30L1* gene consisted of a DNA segment with 4.5 kb of upstream and 2.5 kb of downstream sequences (relative to the transcription unit) that also included another four genes: *CG6443*, *CG17118*, *CG6750* and *Nup170*. To disrupt the *dpy-30L1* gene, the coding sequence of YFP with its genuine stop codon followed by the SV40 polyadenylation/transcription termination sequence, was inserted in frame immediately following the ATG start codon. Insertion of YFP resulted in disruption of *dpy-30L1* reading frame as well as a deletion of 17 aa from the coding region; the truncated Dpy-30L1 sequence was out of frame relative to the ATG-YFP sequence.

The targeting construct for *dpy-30L2* gene contained 3.1 kb of upstream and 3.6 kb of downstream sequences of the gene. Also in this case, the coding sequence of YFP with its stop codon and the SV40 sequence was inserted after the ATG. Insertion of YFP resulted in the disruption of the *dpy-30L2* reading frame and in this case deletion of 40 aa from the coding region, generating the following junction: CCTCAGCCCAACAatgC/CCGGACACCAGTTCCATG, where atg stands for the initiator triplet and slash indicates the junction.

Targeting constructs contained an *I-SceI* cleavage site and were inserted into the pTARG plasmid that contained a multiple cloning site, an *I-CreI* recognition site, a *mini-white* gene, two *loxP* sites, and two FLP recombinase target sites to release a circular episome for gene targeting [49]. Targeting was performed by a procedure essentially corresponding to that described by [50,51,36]. By screening a total of 23 000 flies, we recorded five independent events for *dpy-30L1* (i.e., a frequency of one event in 4600 flies) and two independent events from 6 000 screened flies for *dpy-30L2*, respectively (a frequency of one event in 3 000 flies). The reduction efficiency of the two tandem copies to the mutant was 32% for *dpy-30L1* and 20% for dpy-*30L2*. Verification of knockout copies was done using PCR with primers that yielded a different product size in the case of the mutant copy, namely, 1.4 kb vs. 512 bp (wt) for *dpy-30L1* and 1.1 kb vs. 200 bp (wt) for dpy-*30L2*. Sequencing of the fragment confirmed the expected deletion junction: CACATTGCCatgGAGGC/GCTGGCAAGGAGCCAAATG (atg, initiator triplet; slash, junction).

Furthermore, S1 nuclease protection assay revealed a complete absence of genuine mRNA from the two mutated genes.

Genomic rescue

The rescue construct of *dpy-30L2* contained 3.5 kb of upstream sequence and 4.4 kb of downstream sequence relative to the transcription unit, whereby the start of *dpy-30L2* overlaps with the end (400 bp) of the first exon of another gene, namely, *CG1136*.

The cDNA rescue constructs of mammalian orthologs all contained the 3.5 kb upstream region, the leader of the *dpy-30L2* transcript and 4.4 kb downstream sequence.

Expression pattern determination by fluorescent protein reporter

Three different transgenic lines that carried knockout constructs (described above) were used to determine the promoter activity of the genes in different tissues of larvae and flies. Pictures were taken with a Zeiss Axiocam.

Preparation of nuclear extracts for EMSA

Drosophila Schneider S2 cells were transiently transfected with the respective constructs and collected 72 hours later. Electrophoretic mobility shift assays (EMSAs) were performed as described by Radtke *et al.* [17] and Zhang *et al.* [18]. Binding reactions were performed by incubating 25 fmoles of [γ-^{32}P]ATP end-labeled, 31-bp-long double stranded DNA oligonucleotides containing the core MRE consensus sequence (MRE-s), TGCACAC, with nuclear extracts prepared according to Schreiber *et al.* [52]. For competition experiments, 5 pmoles of unlabeled oligonucleotides were added to the reaction mixture prior to the addition of the extracts. The MRE-s oligonucleotide used for EMSA is as follows:

5'-CGAGGGAGCTCTGCACACGGCCCGAAAAGTG-3'
and

3'-TCGAGCTCCCTCGAGACGTGTGCCGGGCTTT-TCACAGCT-5.

Dpy-30L2 and male sterility phenotype

Constructs H2AvD-GFP, Protamine B-eGFP and Mst77F-eGFP-eGFP, used to verify loss of histones with concomitant appearance of protamines and Mst77F during nuclear shaping and chromatin condensation of sperm, are described in Jayaramaiah Raja and Renkawitz-Pohl [40].

Authors' contributions

AV did most of the experiments. LA performed control dissections of sperm motility and did essentially all the manuscript handling. DE initiated the study and helped in designing constructs for targeted gene disruption. SJR and RR–P carried out fertilization and followed the fate of sperm in females, MS–Z analyzed the expression of dpy-30L2-YFP and sperm motility in females. The project was

conceived and jointly supervised by WS and OG. All authors read and approved the final manuscript.

Acknowledgements

We are grateful to Drs. Antonia Manova and Bruno Schmid for technical support, Hasmik Yepiskoposyan, Anand Selvaraj and Dominik Steiger for valuable discussions, to Alisa Davis and George Hausmann for critical reading of the manuscript. This work was supported by the Schweizerischer Nationalfonds, the Kanton of Zürich and by a grant from the project "Controlled Gene Integration" (LSHG-CT-2003-503303) of the European Union to W.S. and by the Deutsche Forschungsgemeinschaft Re628/12-2, Forschergruppe Chromatin mediated biological decisions to R. R.-P.

References

1. Andrews GK, Lee DK, Ravindra R, Lichtlen P, Sirito M, Sawadogo M, Schaffner W: **The transcription factors MTF-1 and USF1 cooperate to regulate mouse metallothionein-I expression in response to the essential metal zinc in visceral endoderm cells during early development.** *Embo J* 2001, **20(5):**1114-1122.
2. Majumder S, Ghoshal K, Gronostajski RM, Jacob ST: **Downregulation of constitutive and heavy metal-induced metallothionein-I expression by nuclear factor I.** *Gene Expr* 2001, **9(4–5):**203-215.
3. Wang Y, Lorenzi I, Georgiev O, Schaffner W: **Metal-responsive transcription factor-1 (MTF-1) selects different types of metal response elements at low vs. high zinc concentration.** *Biol Chem* 2004, **385(7):**623-632.
4. Ogra Y, Suzuki K, Gong P, Otsuka F, Koizumi S: **Negative regulatory role of Sp1 in metal responsive element-mediated transcriptional activation.** *J Biol Chem* 2001, **276(19):**16534-16539.
5. Cramer M, Nagy I, Murphy BJ, Gassmann M, Hottiger MO, Georgiev O, Schaffner W: **NF-kappaB contributes to transcription of placenta growth factor and interacts with metal responsive transcription factor-1 in hypoxic human cells.** *Biol Chem* 2005, **386(9):**865-72.
6. Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, Li Y, Hao YL, Ooi CE, Godwin B, Vitols E, *et al.*: **A protein interaction map of *Drosophila melanogaster*.** *Science* 2003, **302(5651):**1727-1736.
7. **FlyBase: A Database of *Drosophila* Genes & Genomes** [http://www.flybase.org]
8. Hsu DR, Meyer BJ: **The dpy-30 gene encodes an essential component of the *Caenorhabditis elegans* dosage compensation machinery.** *Genetics* 1994, **137(4):**999-1018.
9. Hsu DR, Chuang PT, Meyer BJ: **DPY-30, a nuclear protein essential early in embryogenesis for *Caenorhabditis elegans* dosage compensation.** *Development* 1995, **121(10):**3323-3334.
10. Roguev A, Schaft D, Shevchenko A, Pijnappel WW, Wilm M, Aasland R, Stewart AF: **The *Saccharomyces cerevisiae* SET1 complex includes an Ash2 homologue and methylates histone 3 lysine 4.** *Embo J* 2001, **20(24):**7137-7148.
11. Nislow C, Ray E, Pillus L: **SET1, a yeast member of the trithorax family, functions in transcriptional silencing and diverse cellular processes.** *Mol Biol Cell* 1997, **8(12):**2421-2436.
12. Corda Y, Schramke V, Longhese MP, Smokvina T, Paciotti V, Brevet V, Gilson E, Geli V: **Interaction between SET1p and checkpoint protein Mec3p in DNA repair and telomere functions.** *Nat Genet* 1999, **21(2):**204-208.
13. Schramke V, Neecke H, Brevet V, Corda Y, Lucchini G, Longhese MP, Gilson E, Geli V: **The SET1Delta mutation unveils a novel signaling pathway relayed by the Rad53-dependent hyperphosphorylation of replication protein A that leads to transcriptional activation of repair genes.** *Genes Dev* 2001, **15(14):**1845-1858.
14. Newlon MG, Roy M, Morikis D, Hausken ZE, Coghlan V, Scott JD, Jennings PA: **The molecular basis for protein kinase A anchoring revealed by solution NMR.** *Nat Struct Biol* 1999, **6(3):**222-227.
15. Andrews J, Bouffard GG, Cheadle C, Lu J, Becker KG, Oliver B: **Gene discovery using computational and microarray analysis of transcription in the *Drosophila melanogaster* testis.** *Genome Res* 2000, **10(12):**2030-2043.
16. Westin G, Schaffner W: **A zinc-responsive factor interacts with a metal-regulated enhancer element (MRE) of the mouse metallothionein-I gene.** *Embo J* 1988, **7(12):**3763-3770.
17. Radtke F, Heuchel R, Georgiev O, Hergersberg M, Gariglio M, Dembic Z, Schaffner W: **Cloned transcription factor MTF-1 activates the mouse metallothionein I promoter.** *Embo J* 1993, **12(4):**1355-1362.
18. Zhang B, Egli D, Georgiev O, Schaffner W: **The *Drosophila* homolog of mammalian zinc finger factor MTF-1 activates transcription in response to heavy metals.** *Mol Cell Biol* 2001, **21(14):**4505-4514.
19. Balamurugan K, Egli D, Selvaraj A, Zhang B, Georgiev O, Schaffner W: **Metal-responsive transcription factor (MTF-1) and heavy metal stress response in *Drosophila* and mammalian cells: a functional comparison.** *Biol Chem* 2004, **385(7):**597-603.
20. Gunes C, Heuchel R, Georgiev O, Muller KH, Lichtlen P, Bluthmann H, Marino S, Aguzzi A, Schaffner W: **Embryonic lethality and liver degeneration in mice lacking the metal-responsive transcriptional activator MTF-1.** *Embo J* 1998, **17(10):**2846-2854.
21. Andrews GK: **Cellular zinc sensors: MTF-1 regulation of gene expression.** *Biometals* 2001, **14(3–4):**223-37.
22. Wang Y, Wimmer U, Lichtlen P, Inderbitzin D, Stieger B, Meier PJ, Hunziker L, Stallmach T, Forrer R, Rulicke T, Georgiev O, Schaffner W: **Metal-responsive transcription factor-1 (MTF-1) is essential for embryonic liver development and heavy metal detoxification in the adult liver.** *FASEB J* 2004, **18(10):**1071-9.
23. Lichtlen P, Schaffner : **Putting its fingers on stressful situations: the heavy metal-regulatory transcription factor MTF-1.** *Bioessays* 2001, **23(11):**1010-1017.
24. Giedroc DP, Chen X, Apuy JL: **Metal response element (MRE)-binding transcription factor-1 (MTF-1): structure, function, and regulation.** *Antioxid Redox Signal* 2001, **3(4):**577-96. Review
25. Andrews GK: **Regulation of metallothionein gene expression by oxidative stress and metal ions.** *Biochem Pharmacol* 2000, **59(1):**95-104.
26. Kagi JH: **Overview of metallothionein.** *Methods Enzymol* 1991, **205:**613-626.
27. Palmiter RD: **The elusive function of metallothioneins.** *Proc Natl Acad Sci USA* 1998, **95(15):**8428-8430.
28. Egli D, Selvaraj A, Yepiskoposyan H, Zhang B, Hafen E, Georgiev O, Schaffner W: **Knockout of 'metal-responsive transcription factor' MTF-1 in *Drosophila* by homologous recombination reveals its central role in heavy metal homeostasis.** *Embo J* 2003, **22(1):**100-108.
29. Selvaraj A, Balamurugan K, Yepiskoposyan H, Zhou H, Egli D, Georgiev O, Thiele DJ, Schaffner W: **Metal-responsive transcription factor (MTF-1) handles both extremes, copper load and copper starvation, by activating different genes.** *Genes Dev* 2005, **19(8):**891-896.
30. Wimmer U, Wang Y, Georgiev O, Schaffner W: **Two major branches of anti-cadmium defense in the mouse: MTF-1/metallothioneins and glutathione.** *Nucleic Acids Res* 2005, **33(18):**5715-5727.
31. Yepiskoposyan H, Egli D, Fergestad T, Selvaraj A, Treiber C, Multhaup G, Georgiev O, Schaffner W: **Transcriptome response to heavy metal stress in *Drosophila* reveals a new zinc transporter that confers resistance to zinc.** *Nucleic Acids Res* 2006, **34(17):**4866-4877.
32. Schneider I: **Cell lines derived from late embryonic stages of *Drosophila melanogaster*.** *J Embryol Exp Morphol* 1972, **27(2):**353-365.
33. **FlyTED: the Drosophila Testis Gene Expression Database** [http://www.fly-ted.org]
34. **FlyAtlas: the Drosophila adult gene expression atlas** [http://www.flyatlas.org]
35. Chintapalli VR, Wang J, Dow JAT: **Using FlyAtlas to identify better *Drosophila* models of human disease.** *Nature Genetics* **39:**715-720.
36. Rong YS, Titen SW, Xie HB, Golic MM, Bastiani M, Bandyopadhyay P, Olivera BM, Brodsky M, Rubin GM, Golic KG: **Targeted mutagenesis by homologous recombination in *D. melanogaster*.** *Genes Dev* 2002, **16(12):**1568-1581.
37. Miller T, Krogan NJ, Dover J, Erdjument-Bromage H, Tempst P, Johnston M, Greenblatt JF, Shilatifard A: **COMPASS: a complex of proteins associated with a trithorax-related SET domain protein.** *Proc Natl Acad Sci USA* 2001, **98(23):**12902-12907.

38. Nagy PL, Griesenbeck J, Kornberg RD, Cleary ML: **A trithorax-group complex purified from *Saccharomyces cerevisiae* is required for methylation of histone H3.** *Proc Natl Acad Sci USA* 2002, **99(1):**90-94.
39. Russell SR, Kaiser K: ***Drosophila melanogaster* male germ line-specific transcripts with autosomal and Y-linked genes.** *Genetics* 1993, **134(1):**293-308.
40. Jayaramaiah Raja S, Renkawitz-Pohl R: **Replacement by *Drosophila melanogaster* protamines and Mst77F of histones during chromatin condensation in late spermatids and role of sesame in the removal of these proteins from the male pronucleus.** *Mol Cell Biol* 2005, **25(14):**6165-6177.
41. Rong YS: **Gene targeting by homologous recombination: a powerful addition to the genetic arsenal for *Drosophila* geneticists.** *Biochem Biophys Res Commun* 2002, **297(1):**1-5.
42. Briggs SD, Bryk M, Strahl BD, Cheung WL, Davie JK, Dent SY, Winston F, Allis CD: **Histone H3 lysine 4 methylation is mediated by SET1 and required for cell growth and rDNA silencing in *Saccharomyces cerevisiae*.** *Genes Dev* 2001, **15(24):**3286-3295.
43. Ercan S, Giresi PG, Whittle CM, Zhang X, Green RD, Lieb JD: **X chromosome repression by localization of the *C. elegans* dosage compensation machinery to sites of transcription initiation.** *Nat Genet* 2007, **39(3):**403-408.
44. Lüscher-Firzlaff J, Gawlista I, Vervoorts J, Kapelle K, Braunschweig T, Walsemann G, Rodgarkia-Schamberger C, Schuchlautz H, Dreschers S, Kremmer E, Lilischkis R, Cerni C, Wellmann A, Lüscher B: **The human trithorax protein hASH2 functions as an oncoprotein.** *Cancer Res* **68(3):**749-58. 2008 Feb 1;
45. Milne TA, Briggs SD, Brock HW, Martin ME, Gibbs D, Allis CD, Hess JL: **MLL targets SET domain methyltransferase activity to Hox gene promoters.** *Molecular cell* 2002, **10(5):**1107-1117.
46. Issaeva I, Zonis Y, Rozovskaia T, Orlovsky K, Croce CM, Nakamura T, Mazo A, Eisenbach L, Canaani E: **Knockdown of ALR (MLL2) reveals ALR target genes and leads to alterations in cell adhesion and growth.** *Mol Cell Biol* 2007, **27(5):**1889-1903.
47. Fuller MT: **Genetic control of cell proliferation and differentiation in *Drosophila* spermatogenesis.** *Seminars in cell & developmental biology* 1998, **9(4):**433-444.
48. Okazaki Y, Furuno M, Kasukawa T, Adachi J, Bono H, Kondo S, Nikaido I, Osato N, Saito R, Suzuki H, *et al.*: **Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs.** *Nature* 2002, **420(6915):**563-573.
49. Egli D, Yepiskoposyan H, Selvaraj A, Balamurugan K, Rajaram R, Simons A, Multhaup G, Mettler S, Vardanyan A, Georgiev O, Schaffner W: **A family knockout of all four *Drosophila* metallothioneins reveals a central role in copper homeostasis and detoxification.** *Mol Cell Biol* 2006, **26(6):**2286-2296.
50. Rong YS, Golic KG: **Gene targeting by homologous recombination in *Drosophila*.** *Science* 2000, **288(5473):**2013-2018.
51. Rong YS, Golic KG: **A targeted gene knockout in *Drosophila*.** *Genetics* 2001, **157(3):**1307-1312.
52. Schreiber E, Matthias P, Müller MM, Schaffner W: **Rapid detection of octamer binding proteins with 'mini-extracts', prepared from a small number of cells.** *Nucleic Acids Res* **17(15):**6419. 1989 August 11;
53. Wood KV, Lam YA, McElroy WD: **Introduction to beetle luciferases and their applications.** *J Biolumin Chemilumin* 1989, **4(1):**289-301.
54. McKeown M, Belote JM, Boggs RT: **Ectopic expression of the female transformer gene product leads to female differentiation of chromosomally male *Drosophila*.** *Cell* 1988, **53(6):**887-895.

Acknowledgements

I would like to express my deep gratitude to my thesis supervisor Walter Schaffner for being a wonderful mentor and just a bighearted and interesting person. Thanks to Walter, I was encouraged to work on various projects and obtain practical skills and develop new thinking abilities. As an indispensable part of my gratefulness to him are also the non-scientific activities like hikes in Swiss Alps, games and competitions, cognac aperos and Mandelbärli. I have learnt a lot about "Mother Nature" and Switzerland from Walter!

There are no words to describe what Dr. Georgiev (or simply Oleg) meant for me during PhD and beyond. Our endless discussions covered many important topics, from molecular cloning to Bulgarian yogurt recipe. He was always there for me in all the practical and "cloning strategy" issues. Oleg was also my great teacher for Goodminton! Next to Oleg I am also grateful to Dr. Antonia Manova ("Toni") for her essential technical support and nice morning smile and caring nature.

My life in Switzerland during PhD years was nicely complemented with my Armenian friends: Artur, Hasmik (for our outstanding trips and memorable evenings in Bern), Alla (for her help in the beginning of my PhD and thereafter), Ruzanna (for listening to my complaints and cheering me up) and Armine and Margarita (for nice presents from Armenia and joy and laughter). There are even more Armenian friends I am thankful to, namely Artash, Vahe, Karen, Arpi who were also my flatmates in Zurich and nice company for evenings.

A huge "thank you" goes certainly to Viola and Uschi (officially Dr. Günther and Dr. Lindert, respectively). Except a lot of lab skills they taught me, there is a long list of the things I am thankful for: German classes, nice travel company, Christmas markets, movie evenings, healthy lunch breaks, etc. Uschi was the person who showed me the S1 mapping, one of the key experiments. We passed through many things with Viola, even through fire (Moscow, 2010) and I am thankful to her for being always very supportive and kind.

A lot of thanks to Dominique for nice surprises on my desk or on the door; what a wonderful friend, flatmate and colleague.

A special gratitude to Dominik Steiger for *Drosophila* genetics and tango!

I am also thankful to former and current colleagues and friends: Chetak (for endless talks in the IMB corridors), Patrizia, Haiqing, Kurt, Till (for ice-cream breaks and jokes), Prasuna, Sabari, Sreehari, Corinna, Natalia, Liz, Michi, Michelle, Julia, Irma, Maria, Sonya, Susanna Bachmann, Bernhard Dichtl (for his enormous help with yeast experiments), Erich Frei and Markus Noll (for critical and helpful comments), Werner Boll (for the help with microscopes), Michi Daube (for TGIF).

A very special “thank you” to Shavarsh, whom I met three days after I started writing the dissertation, and he filled the following time with flowers, love and inspiration.

Last, but most importantly, I would like to thank my family: parents Armen and Lyudmila and sisters Anahit and Sirvard and my cousin Arusyak. I greatly appreciate their support, love and care during my entire life.

Curriculum Vitae

Name Surname Lilit Atanesyan

Date and Place of Birth 17 November 1984, Yerevan, Armenia

Nationality Armenian

Education

July 2007 – to date
PhD student, Institute of Molecular Life Sciences,
University of Zurich, Zurich, Switzerland.
Life Science Zurich Graduate School (ETH Zurich, University of Zurich).

September 2005 – June 2007, Master's degree.
Department of Biophysics, Faculty of Biology,
Yerevan State University, Yerevan, Armenia.
Master's thesis title:
Implication of Poly(ADP-ribosyl)ation in Internucleosomal Chromatin Cleavage.

September 2001 – July 2005, Bachelor's Degree, Honor's diploma.
Department of Biophysics, Faculty of Biology,
Yerevan State University, Yerevan, Armenia.
Bachelor's thesis title:
Age-related and Hydrocortisone-induced Characteristics of Oligonucleosomal DNA Fragmentation in Thymus of Rats.

Title of Dissertation

"Polyglutamine-mediated and Heavy Metal-induced Transcription from Yeast to Humans"

Publications

Steiger D, Fetchko M, Vardanyan A, Atanesyan L, Steiner K, Turski ML, Thiele DJ, Georgiev O, Schaffner W. The Drosophila copper transporter Ctr1C functions in male fertility. J Biol Chem. **2010** May 28;285(22):17089-97.

Bhardwaj N, Strickland AD, Ahmad F, Atanesyan L, West K, Lloyd DM. A comparative histological evaluation of the ablations produced by microwave, cryotherapy and radiofrequency in the liver. Pathology. **2009** Feb;41(2):168-72.

Vardanyan A, Atanesyan L, Egli D, Raja SJ, Steinmann-Zwicky M, Renkawitz-Pohl R, Georgiev O, Schaffner W. Dumpy-30 family members as determinants of male fertility and interaction partners of metal- responsive transcription factor 1 (MTF-1) in Drosophila. BMC Dev Biol. **2008** Jun 27;8:68

Melnikova-Sharova M, Torgomyan T, Margaryan A, Galstyan A, Atanesyan L, Sargsyan E. Insulin action on oligonucleosomal DNA fragmentation in thymocyte nuclei. *Journal of Scientific Notice*, Yerevan State University. **2008** (1):96-101.